evolution

BOTANY REFERENCE NOTES

Paper – II

Genetics - Part 1 and Part 2

Topics covered:

Part 1: Transmission Genetics

Part 2: Molecular Genetics

evolution
Priyanka K. 8923876687

evolution
A Centre of Excellence for Civil Service Examination Guidance
B 11, Second Floor, Commercial Complex, Batra Cinema Road
Next to ICICI A.T.M., Dr. Mukherjee Nagar, DELHI - 110 009
Phone: 011 - 32974645, 32974651, 47092329

https://t.me/Upsc_Pdf

evolution

A Centre of Excellence for Civil Service Examination Guidance

B 11, Second Floor, Commercial Complex, Batra Cinema Road
Next to ICICI A.T.M., Dr. Mukherjee Nagar, DELHI - 110 009
Phone: 011 - 32974645, 32974651, 47092329

Genetics - I

Priyanka

© Copyright Evolution Educare Pvt. Ltd.

All rights are owned by or controlled by Evolution Educare Pvt. Ltd. The content of this book cannot be copied, reproduced, republished, posted, broadcast or transmitted in any way without first obtaining Evolution Educare Pvt. Ltd.'s written permission.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Detailed Table of Contents

Priyanka. 8802387667

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Priyanka K. 8802387667

Direct PDF https://t.me/CivilServices_UPSC

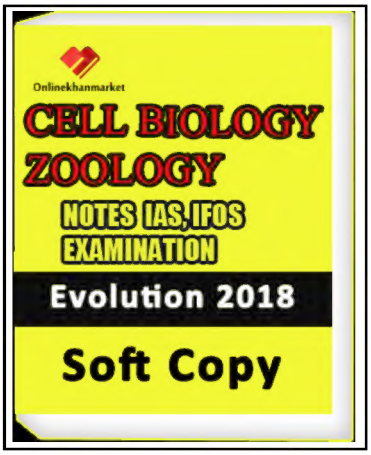
https://t.me/Upsc_Pdf

Priyanka K. 8802387667

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Priyanka K. 8802387667
UPSC Pdf.com



Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Prescribed syllabus of Genetics

For Civil Service Examination

Covered in Genetics Book – I (Transmission Genetics)

Development of genetics; Gene versus allele concepts (Pseudoalleles); Quantitative genetics and multiple factors; Incomplete dominance, polygenic inheritance, multiple alleles; Linkage and crossing over; Methods of gene mapping, including molecular maps (idea of mapping function); Sex chromosomes and sex-linked inheritance, sex determination and molecular basis of sex differentiation; Cytoplasmic inheritance and cytoplasmic genes (including genetics of male sterility).

Covered in Genetics Book – II (Molecular Genetics)

Structure and synthesis of nucleic acids and proteins;Genetic code and regulation of gene expression; Gene silencing; Multigene families; Organic evolution – evidences, mechanism and theories. Role of RNA in origin and evolution.

Mutations (biochemical and molecular basis)

For Indian Forest Service Examination

Covered in Genetics Book – I (Transmission Genetics)

Development of genetics, and gene versus allele concepts (Pseudoalleles). Quantitative genetics and multiple factors. Linkage and crossing over–methods of gene mapping including molecular maps (idea of mapping function). Sex chromosomes and sexlinked inheritance, sex determination and molecular basis of sex differentiation. Cytoplasmic inheritance and cytoplasmic genes (including genetics of male sterility).

Covered in Genetics Book – II (Molecular Genetics)

Structure and synthesis of nucleic acids and protines. Genetic code and regulation of gene expression. Multigene families. Organic evolution-evidences, mechanism and theories. Role of RNA in origin and evolution.

Mutation (biochemical and molecular basis).

Prions and prion hypothesis. (Already covered in the book on MICROBIOLOGY)

Priyanka.8802387697

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 1: Development of genetics

Introduction to the field of genetics

Genetics is the study of the transmission, function and behavior of genes. The modern science of genetics influences many aspects of human life, including food & nutrition, disease diagnostics, prevention and treatment, industrial processes, agriculture, environmental remediation and several others. Molecular genetics advances in recent times have enabled scientists to alter a microbe, plant or animal to make it more useful.

The development of genetics as a discipline

The field of genetics has evolved rather exponentially during the last about one hundred and fifty years. In fact, the development in this field has been so fast paced during the recent decades that many people call it the period of *Genetic Revolution*. However, in totality the field of genetics has developed because of large number of scientific discoveries and breakthroughs, which have taken place over the years. The development of genetics can be summarized as follows.

Early Views of Heredity

In ancient times, people understood some basic rules of heredity and used this knowledge to breed domestic animals and crops. By about 5000 bc, for example, people in different parts of the world had begun applying selective breeding techniques to grow new plant varieties, including types of wheat, maize, rice, and date palms, that had never existed in the wild.

Ancient people understood that the rules of inheritance also applied to humans. The ancient Greeks were particularly interested in human heredity and evolution. Greek scientists and philosophers hotly debated whether a male or female parent contributed more to an offspring. In the 4th century bc, Aristotle speculated that acquired characteristics, such as a scar that was incurred during life, could be passed on to offspring. He also believed in a widely held theory known as *pangenesis*. This theory proposed that particles in the body, called gemmules, reside in the limbs and organs. The gemmules become imprinted with any changes acquired by the body, such as muscle development from exercise. The gemmules then move to the reproductive cells and transfer information about the body's alterations to these cells. The reproductive cells transmit the acquired traits to offspring through particles called pangenes.

The theories about the inheritance of acquired characteristics and pangenesis persisted until the middle of the 19th century. French zoologist Jean-Baptiste Lamarck formalized the *theory of acquired characteristics* in his treatise *Philosophie Zoologique* (1809). Lamarck proposed that organisms evolve by responding to changes in their environment. When organisms undergo a change in order to adjust to their environment, that change acts as a trait that can be passed on to offspring.

There was also the concept of *blending inheritance*: the idea that individuals inherit a smooth blend of traits from their parents.

Influences of Mendel & the development of classical genetics

The modern science of genetics traces its roots to Gregor Johann Mendel, a German-Czech monk and scientist who studied the nature of inheritance in plants. In his paper "*Versuche über Pflanzenhybriden*" ("Experiments on Plant Hybridization"), presented in 1865 to the *Naturforschender Verein* (Society for Research in Nature) in Brünn, Mendel traced the inheritance patterns of certain traits in pea plants and described them mathematically. Although this pattern of inheritance could only be observed for a few traits, Mendel's work suggested that

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

heredity was particulate, not acquired, and that the inheritance patterns of many traits could be explained through simple rules and ratios.

Mendel conducted a long series of experiments on pea plants during the 1850s and 1860s. Mendel crossbred plants that expressed differing traits, such as height and flower color. His conclusions from these experiments helped him formulate a comprehensive theory of how such traits pass from one generation to another. In his studies, Mendel recognized that characteristics were inherited as discrete units, and that each of these was inherited independently of the others. He speculated that each parent has pairs of these units but passes only one to an offspring. He also noted that certain forms of one trait were always dominant over others. Today, the units that Mendel described are known as genes.

The importance of Mendel's work did not gain wide understanding until the 1890s. In the year 1900, however, Dutch botanist Hugo Marie de Vries, German botanist Karl Correns, and Austrian botanist Erich Tschermak independently rediscovered the monk's works and verified his conclusions.

After the rediscovery of Mendel's work, scientists tried to determine which molecules in the cell were responsible for inheritance. Advances in cytology, the science of the structure and function of cells, enabled scientists to more deeply appreciate Mendel's work.

In 1902 American biologist Walter S. Sutton and German cell biologist Theodor Boveri separately noted the parallels between Mendel's units and chromosomes. The demonstration of the *chromosomal basis of inheritance* gave rise to the modern science of genetics. The term *genetics* itself was coined in 1905 by British biologist William Bateson. The terms *gene* and *genotype* were contributed in 1909 by German scientist Wilhelm Johannsen.

In 1905 American biologists Edmund B. Wilson and Nettie Stevens independently discovered and identified the sex chromosomes. Wilson discovered the X chromosome in a butterfly, and Stevens discovered the Y chromosome in a beetle. The discoveries of the X and Y chromosomes helped scientists begin to unravel new patterns of inheritance.

Foremost among this research was the work of American biologist Thomas Hunt Morgan on *Drosophila*. In 1910, Thomas Hunt Morgan argued that genes are on chromosomes, based on observations of a sex-linked white-eye mutation in *Drosophila*. In 1913, his student Alfred Sturtevant used the phenomenon of genetic linkage to show that genes are arranged linearly on the chromosome.

More definitive proof emerged in the 1930s with work by American geneticists Harriet Creighton and Barbara McClintock. The pair demonstrated gene recombination with experiments on seed color in corn. McClintock later gained notice for her work on transposable elements, large genetic segments that move within a chromosome or even between chromosomes. Her research into these elements, commonly known as jumping genes, earned McClintock the 1983 Nobel Prize in physiology or medicine.

The Development of Biochemical and Molecular Genetics

While cytologists and geneticists were studying the properties and location of genes on chromosomes, other scientists focused their studies on the composition of genes. In 1928 British microbiologist, Frederick Griffith ran a series of experiments on two strains of bacteria, one that kills mice and another that is harmless to them. When Griffith injected mice with killed cells of the virulent bacteria, all of the mice survived. However, in a second trial, when Griffith injected a combination of dead virulent bacteria and live "harmless" bacteria, the mice all died. He

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

concluded that something in the dead virulent cells "transformed" the hereditary material of normally harmless bacteria so that they became killers. Most scientists at the time theorized that the transforming factor was composed of a protein.

The real identity of the transforming factor in this experiment was not identified until 1944, when American geneticists Oswald Avery, Colin MacLeod, and Maclyn McCarty revisited Griffith's research. After isolating different molecular components from dead bacterial cells, Avery and his colleagues determined that DNA was the agent that transformed the live harmless bacteria into killers.

Despite a growing body of evidence about the function of DNA, many scientists were not ready to reject proteins as the hereditary material. The debate was largely quieted in 1952 by American geneticists Alfred Hershey and Martha Chase. Hershey and Chase showed that when a type of virus called a bacteriophage infects a bacterium, it is the virus's DNA—not protein—that enters the bacterium to cause infection. Their studies confirmed that DNA contained the virus's genetic information, which triggered viral replication within the bacteria.

The experiments of Hershey and Chase convinced most scientists that DNA was the molecule of heredity, but many questions about the structure and mechanisms of DNA remained. In the early 1950s researchers began to apply techniques of X-ray diffraction to learn about the basic structure of DNA. X-ray diffraction can determine molecular structures by measuring patterns of scattered X rays after they pass through a crystalline substance. British physical chemist Rosalind Franklin and British biophysicist Maurice Wilkins used X-ray diffraction to obtain DNA images of unprecedented clarity.

Yet the exact three-dimensional structure of DNA remained unclear. The groundbreaking work of American biochemist James Watson and British biophysicist Francis Crick solved that mystery. In 1953 the two proposed a model of DNA that is still accepted today: A double helix molecule formed by two chains, each composed of alternating sugar and phosphate groups, connected by nitrogenous bases. Watson and Crick (along with Wilkins) were awarded the 1962 Nobel Prize in physiology or medicine for their discoveries.

Watson and Crick speculated that the structure of DNA provided some obvious clues about how the molecule could replicate itself. They proposed a replication model in which each strand of DNA serves as a template for making exact copies. This model of replication, called semi-conservative replication, was demonstrated in 1958 by American molecular biologists Matthew Meselson and Franklin Stahl. Their experiments demonstrated the mechanisms of replication by tracking DNA containing a heavy nitrogen isotope through a series of replications.

With DNA's structure and replication mechanisms largely solved, scientists turned their attention to identifying the genetic code—learning how a gene's nucleotide sequence determines what type of protein is made. In the late 1950s, South African geneticist Sydney Brenner and other scientists confirmed that RNA acted as an intermediary between DNA and protein production. Researchers still were uncertain how the sequence of nucleotides in DNA corresponded to the production of specific amino acids. In 1961 Crick and Brenner determined that groups of three nucleotides, now known as codons, code for the 20 amino acids that form the foundation of proteins.

The exact relationship between codons and amino acids was clarified after several important discoveries. American biochemists Marshall Nirenberg and J. Heinrich Matthaei synthesized repeated nucleotide sequences that led to the production of repeated single amino acids. They identified how certain codon combinations code for a specific amino acid. A process developed by

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

American geneticist Har Gobind Khorana helped scientists create a "dictionary" of codons that defined specific amino acids, thus resolving the remaining ambiguities in the genetic code. Only 12 years after the structure of DNA was deduced, the genetic code was solved.

The Era of Molecular Manipulation of Genes

After scientists had unraveled the structure and replication mechanisms of DNA, many felt that the major discoveries of genetic research were resolved. They predicted that the only task left in genetics was to sort out the molecular details of how genes work. But in the process of studying gene function, researchers developed powerful new molecular techniques, enabling them to analyze and manipulate genes with a speed and precision never before possible.

A number of discoveries made during the 1960s and 1970s shed light on how distinct fragments of DNA could be isolated. The work of Swiss molecular biologist Werner Arber focused on specialized enzymes that digest, or "restrict," the DNA of viruses infecting bacteria. These enzymes were subsequently dubbed restriction enzymes. In the following decade, scientists learned that restriction enzymes could also act like molecular scissors to cut DNA. In 1970 American molecular biologist Hamilton Smith and colleagues determined that restriction enzymes could cleave DNA molecules at precise and predictable locations. Hamilton concluded that the enzymes were able to recognize specific nucleotide sequences.

Scientists quickly realized that restriction enzymes could be used in the laboratory to manipulate DNA. In 1973 American biochemist Herb Boyer used restriction enzymes to produce a DNA molecule with genetic material from two different sources. This splicing technique is now known as recombinant DNA. Boyer inserted foreign genes into plasmids and observed that the plasmids could replicate to make many copies of the inserted genes. In subsequent experiments, Boyer, American biochemist Stanley Cohen, and other researchers demonstrated that inserting a recombinant DNA molecule into a host bacteria cell would lead to extremely rapid replication and the production of many identical copies of the recombinant DNA. This process, known as cloning, gave scientists the power to make many copies of desired DNA for molecular study.

The speed and efficiency of DNA cloning were vastly improved in the 1980s with the invention of polymerase chain reaction (PCR). Developed by American biochemist Kary Mullis, PCR enables scientists to produce large amounts of DNA sequences in a test tube. In a matter of hours, the process can produce millions of cloned DNA molecules.

Yet all of the advances in isolating and replicating DNA would not be possible or be of much use if researchers could not determine the nucleotide sequence of genetic material. In the late 1970s and early 1980s, British biochemist Frederick Sanger and his associates developed DNA sequencing techniques. Sanger's methods, which used special compounds called dideoxy nucleotides, rapidly yielded the exact nucleotide sequence of a desired sample. With the use of automated equipment, the new techniques transformed genetic sequencing into a speedy, routine laboratory procedure.

These developments encouraged the scientists to start the most ambitious project of genetics, which is the *Human Genome Project* (HGP). Human Genome Project was an international scientific research project with a primary goal to determine the sequence of chemical base pairs which make up human DNA and to identify the approximately 25,000 genes of the human genome from both a physical and functional standpoint. The project began in 1990 initially headed by James D. Watson at the U.S. National Institutes of Health. A working draft of the genome was released in 2000 and a complete one in 2003, with further analysis still being published.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 2: Mendel's laws of inheritance

Basic idea of heredity and variation

Heredity is the transmission of characters from one generation to the next, i.e., from parents to their offspring. Because of heredity, the offspring resemble their parents. Heredity is the essence of self-reproduction. It is owing to heredity or self-reproduction that we commonly observe the phenomenon of "like begets like".

Variations are the visible differences between the parents and the offspring, or between two offsprings of the same parents.

An offspring receives all the characters from its parents and yet, an offspring is never an exact copy of its parents. Similarly, no two offsprings of the same parents are identical (exception: identical twins).

In order to understand the principles of inheritance and to discover the reasons for the variations, Mendel began a systematic search during the second half of the nineteenth century. For this, Mendel experimented on garden pea plants and performed various crosses with great precision, care and objectivity. He carefully counted the plants resulting from such crosses and kept statistical records of successive generations with the accuracy of a mathematician.

Gregor Mendel & his life

Gregor Johann Mendel was born on July 22, 1822 in Moravia, Austria. He had his early education in a monastery in Brunn, Austria (now Brno in Czechoslovakia) and later studied science and mathematics at the University of Vienna. He graduated in 1840. Mendel returned to the monastery in Brno as a monk. He worked as a teacher of physics and natural science in a Higher Secondary School of Brno during 1854 to 1868. He was appointed abbot of the monastery in 1868 and held this post untill his death.

Mendel carried out his legendary experiments on garden pea plants in the monastery garden from 1857 to 1865. He had a clear perspective and worked on the experiments with precision and thoroughness. He published his research paper containing his observations and conclusions in 1866 in the annual proceedings of the Natural History Society of Brunn. These conclusions are now known as Mendel's Laws. This work is a classic in biology for its elegance and simplicity and ranks amongst the most outstanding biological contributions of all times. However, unfortunately, this work failed to attract the attention of the biologists of that time. Hence, it went ignored for 35 years. One of the possible reasons for such neglect was the inability of the biologists of that time to understand and appreciate the statistical approach adapted by Mendel. Thus, Mendel was left bitterly disappointed, and died an unrecognized death in 1884.

Sixteen years after Mendel's death, in 1900, Hugo de Vries (Holland), Karl Korrens (Germany) and Von Tschermark (Austria) independently arrived at similar conclusions as those of Mendel. De Vries rediscovered the research paper of Mendel and it was published again in 1901. The experiments on heredity of plants and animals since then have confirmed that Mendel's laws of heredity are applicable to other organisms as well. They form the basis of modern genetics. Hence, Mendel is called *The Father of Genetics.*

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Mendel's Experiments

Selection of the material: Mendel selected garden sweet pea (*Pisum sativum*) for his hybridization experiments for of the following reasons:

1. Plants are annual and easy to cultivate.
2. Peas have many distinct, well-defined and easily observable morphological characteristics (traits).
3. Flowers are bisexual and naturally self-fertilizing, but they can also be easily cross-fertilized.
4. Plants produce a large number of seeds. Hence, the sample size would be larger when ratios of genetic transmission are obtained.
5. The offspring of cross-fertilized plants are fertile.
6. Flowers are sufficiently large for easy emasculation (removal of stamens) and artificial cross-pollination.

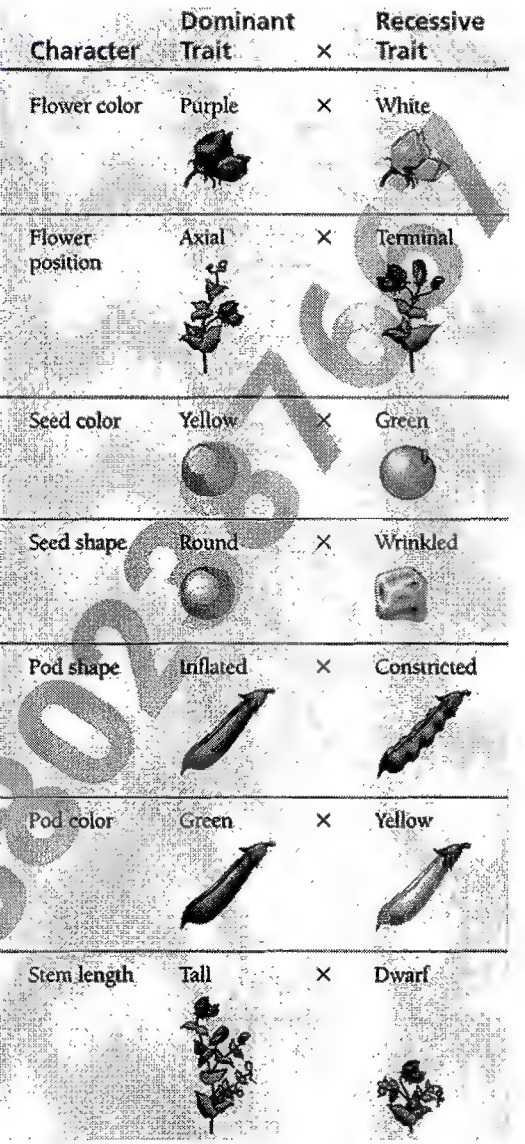


Figure 1: The traits selected by Mendel in *Pisum sativum*

Selection of characters (traits): Mendel selected 14 different varieties of the pea and grouped them into seven pairs. Each pair was considered for a specific trait (characteristic) such as flower color or seed shape or stem length, etc. The two members of each pair showed contrasting forms of the chosen trait, e.g., in a pair selected for stem length, one variety had a tall stem (6-7 feet tall) while the other had a dwarf stem. These seven pairs of contrasting traits are shown in Figure 1.

Nature and procedure of the experiments: Mendel had very clear concepts of what he was doing and what requirements were necessary if he had to arrive at the conclusions accurately and successfully. He realized the necessity of:

1. using pure breeding parent plants;
2. considering only one character at a time during the breeding experiments;
3. always keeping the generations separate; and
4. using statistical and mathematical principles to compute the obtained results properly.

Mendel conducted his experiments in three stages.

Stage-1: It involved selecting a pair of parents with contrasting characters and obtaining each parent plant in pure condition, i.e., breeding true for the characters selected. For example, Mendel ensured that the plant selected for round seeds produced only round seeds on self-fertilization and the plant selected for wrinkled seeds produced only wrinkled seeds. Such pure,

Direct PDF https://t.me/CivilServices_UPSC

true-breeding parents were obtained by Mendel through repeated self-fertilizing, generation after generation.

Stage -2: It involved crossing of the selected parents. Of the pair, one plant was used as the male parent and the other as the female parent. Pollen from the male was dusted on the stigma of the female parent for cross-pollination. Mendel conducted *reciprocal crosses* also. For example: in one cross, the round seeded variety was used as the male parent and the wrinkled seeded variety as the female parent, while in the reciprocal cross of the same parents, the wrinkled variety was used as the male parent and the round seeded variety as the female parent.

Such a cross between two parents representing contrasting forms of a single character is called monohybrid cross and the offspring is called a hybrid. The hybrid represents the first filial generation or F1 generation.

Stage - 3: In the third stage, Mendel allowed each F1 hybrid to self-pollinate and produce the next, i.e., Second Filial generation or F2 generation.

Mendel conducted a similar type of hybridization experiment separately for each of the seven pairs. He meticulously maintained a complete record of the actual number of each type of offspring in every generation (i.e., data of qualitative as well as quantitative results).

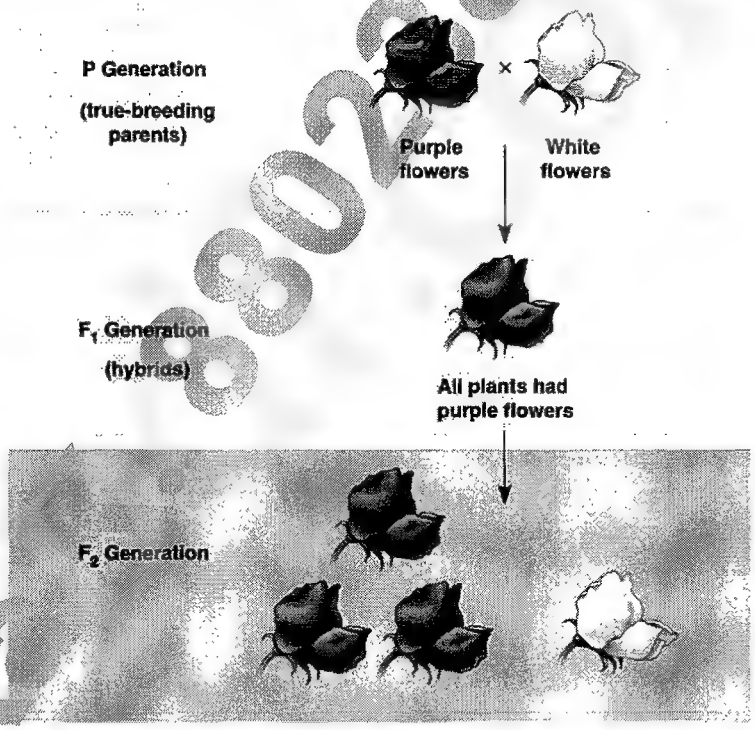


Figure 2: The observations of Mendel in the Monohybrid Crosses

Observations: (Figure 2) Mendel observed that in each of these crosses, all the F1 hybrids resembled only one parent, while the character of the other parent was not seen in the F1 hybrid. For example, in a cross between round and wrinkled type, the F1 were all round seeds only. The wrinkled character was not seen. The character, which appears in the F1 hybrid, was termed **dominant** and the other as **recessive** by Mendel. Thus, in each of the seven pairs, one form is dominant and the other is recessive. The F2 progeny showed presence of both parental forms, which always appeared in the ratio of 3 dominants: 1 recessive. This 3:1 F2 ratio was termed as the ***monohybrid ratio***.

Law of Dominance

Unit characters: Based on his experiments and observations, Mendel concluded that the characters from parents to offsprings are transmitted in the form of some kind of hereditary units or particles called factors (now called genes). These factors determine the characters in the individual. For each factor representing a character there are two alternate forms called alleles. Mendel represented the allelic factors by appropriate alphabets. In each of the seven pairs

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

selected by him, the dominant allele was represented by the capital alphabet and its recessive allele by the same but smaller alphabet (e.g., Round seed = R, Wrinkled seed = r).

Law of dominance: In a hybrid union, the allele, which expresses itself phenotypically, is the dominant allele while the other allele, which fails to express itself phenotypically, is the recessive allele. The hybrid individual shows phenotypically only the dominant character.

The law of dominance is often described as **Mendel's First Law of Inheritance.**

Dominance is seen in various characters in many plants and animals. However, this is not of universal occurrence. There are many cases where dominance is incomplete or absent.

Monohybrid Ratio: The phenotypic ratio of different types of individuals occurring in the F2 generation of the monohybrid cross is called the monohybrid ratio. In the Mendelian monohybrid experiments, this ratio was always 3:1(i.e., 75% is dominant and 25% is recessive) as shown in Figure 3.

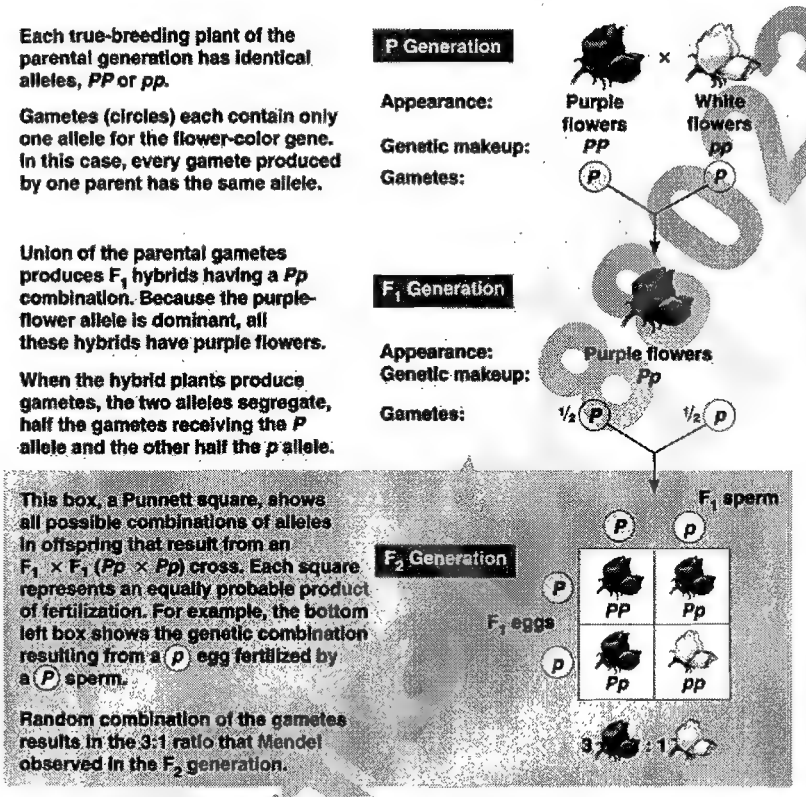


Figure 3: The explanation of the Monohybrid Ratio

According to Mendel, each sexually reproducing diploid organism possesses two 'factors' (genes) for each character; one factor is received (inherited) from male parent and the other factor is inherited from the female parent. These two factors for a particular character are called alleles or allelomorphs. When an offspring receives identical alleles from both parents, it is called homozygous, pure or true breeding for the character. On the other hand, when the offspring receives dissimilar alleles from two parents, it is called heterozygous, impure or a hybrid for that character.

The pure tall is crossed with the pure dwarf parent. According to Mendel, when the diploid individual (having both the alleles/factors) produces gametes, each gamete receives only one of the two factors/alleles of a character. No gamete receives both the alleles of a character. Thus, pure tall parent produces only one type of gametes, i.e. all the gametes possess only (T) factor for *tallness*. Similarly, all gametes produced by pure dwarf are of one type only and possess (t) factor. The fusion of (T) and (t) gametes (fertilization) results in the F1 offspring with (Tt) genotype. It is heterozygous or a hybrid. Its phenotype (external appearance) is tall because the factor for *tallness* (T) is dominant and expresses itself. The factor for *dwarfness* (t) is present in F1 hybrid but, being recessive, does not express itself (remains hidden).

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Mendel allowed hybrids to self-fertilize or inbreed to raise F2 generation. The F1 hybrid has dissimilar alleles (Tt). Therefore, it will produce two types of gametes in equal number i.e. 50% gametes will have (T) factor and remaining 50% will have (t) factor. Since the pea flower is bisexual, it produces both male and female gametes. Thus, the F1 hybrid will produce two types of male gametes (T) and (t) in equal numbers. Similarly, there will be two types of female gametes (T) and (t) in equal numbers. During self fertilization, the fusion between these male and female gametes occurs at random. For example, each type of male gamete has an equal chance to fuse with either (T) or (t) female gametes and vice-versa. This chance fusion, between two types of male and two types of female gametes will produce a maximum of four combinations (genotypes) in the F2 progeny. This is shown in the checker board or Punnet's Square.

These four combinations fall into three categories of the genotypes as follows : 1 (TT), 2 (Tt) and 1 (tt) i.e. 1 *Pure tall*(TT) : 2 *Hybrid tall* 2(Tt) : 1 *Pure dwarf* (tt)

This is called 1:2:1 genotypic ratio of a monohybrid cross. However, phenotypically, the progeny shows 3 Tall and 1 Dwarf individuals (75% Dominant and 25% recessive characters) or 3:1 ratio. This is called monohybrid ratio or phenotypic ratio of a monohybrid cross (Figure 4).

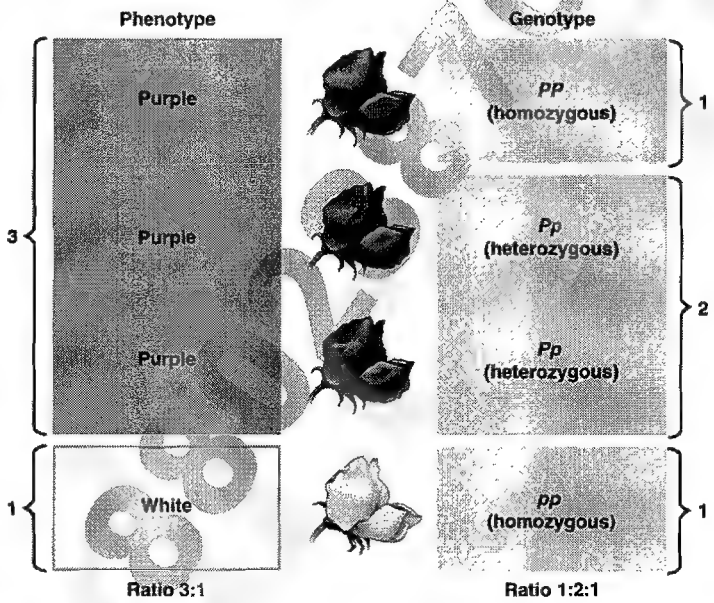


Figure 4: The Genotypic and Phenotypic Ratio

The result indicates that even though the recessive character was not seen in hybrid, it was present there and reappeared in pure form in 25% individuals of the progeny. This result also enabled Mendel to conclude that the two factors (alleles) come together in the hybrid but do not mix or fuse with each other. They simply remain together without diluting or contaminating each other. In other words, factors maintain their purity. As the gamete always receives only one factor (alleles) for a trait, it is always pure for the character. This is called purity of gametes.

Law of Segregation

Mendel performed monohybrid crosses separately for each of the seven pairs of contrasting characters. In each such cross, he got similar results as described for the *Tall x Dwarf* cross. On the basis of the results obtained for the monohybrid crosses, Mendel formulated the **law of segregation**, also called ***Mendel's second law of heredity***.

The law states that when a pair of alleles is brought together in hybrid union, the members of the allelic pair remain together without mixing, diluting or altering each other and separate or segregate from each other when the hybrid forms gametes. Thus, according to this law, when the hybrid tall (Tt) in the above experiment forms gametes, the factors T and t shall separate (segregate) and enter different gametes. As a result, the hybrid shall form two types of gametes; those with (T) factor and an equal number with (t) factor. Since each gamete will be pure for tallness or for dwarfness, the law is also known as the law of the purity of gametes.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

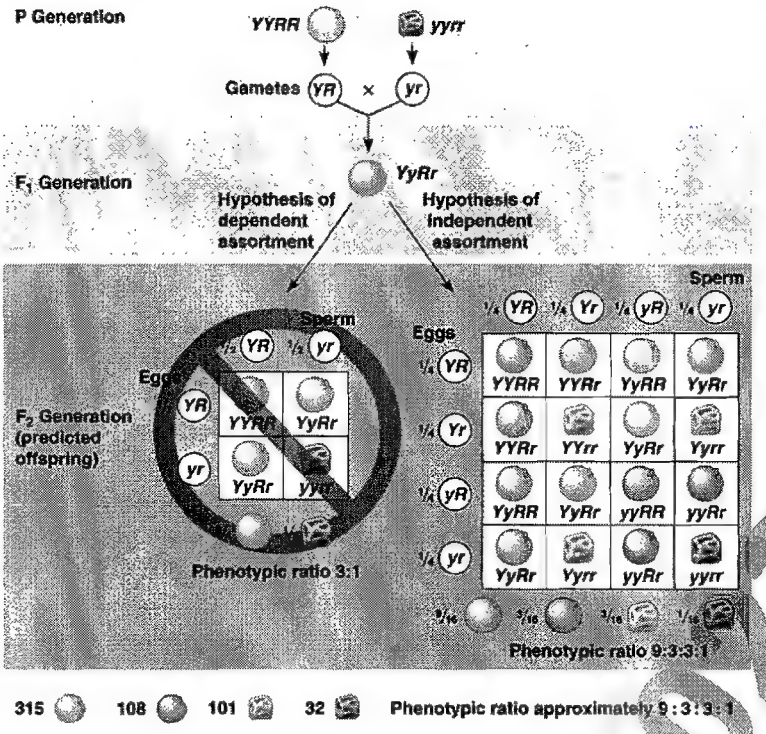


Figure 5: Mendel's dihybrid cross

Dihybrid Ratio: Mendel established the law of segregation through monohybrid crosses involving only one pair of alleles at a time. However, in an organism, so many characters are present together and each character is represented by a pair of alleles. Therefore, Mendel wanted to know, "Whether one pair of alleles affects or influences the inheritance pattern of other pairs of alleles in the organism or, each pair is inherited independently as if in a monohybrid cross?" To find the answers to these questions, Mendel performed dihybrid crosses.

A cross between two pure, true breeding parents in which the inheritance pattern of two allelomorphic pairs is considered (studied) simultaneously is called a dihybrid cross. The phenotypic ratio obtained in the F2 generation of a dihybrid cross is called the dihybrid ratio.

A dihybrid is an individual which is double heterozygous (i.e. heterozygous for two pairs of alleles).

Mendel's dihybrid cross

Mendel considered two characters in the pea plants simultaneously, e.g. cotyledon color (yellow / green) and seed shape (round / wrinkled). He selected one variety of pea which was pure (true breeding) for yellow round seeds and crossed it with another variety pure for green wrinkled seeds. (Figure 5)

All the F_1 of this cross were yellow round seeds (green and wrinkled characters did not appear in F_1 hybrids). Mendel anticipated this because, from the earlier monohybrid experiments he knew that yellow was dominant over green and round was dominant over wrinkled.

Moreover, the reciprocal crosses (interchanging male and female parents) also gave the same results.

Further, when the F_1 dihybrids were self-pollinated or inbred, the F2 generation was always the same.

The analysis of F_2 progeny showed four different kinds of phenotypes. These were (1) Yellow round (2) Yellow Wrinkled (3) Green round and (4) Green Wrinkled in the ratio of 9:3:3:1 respectively. It will be seen that out of these four types, two show the same combinations as the parents whereas the remaining two are new combinations (recombinants).

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

The phenotypic ratio of 9:3:3:1 in the F_2 progeny of a dihybrid cross is called the dihybrid ratio. Same results were obtained by Mendel using other pairs of alleles in different combinations.

On the basis of these experiments and their results, Mendel formulated the law of independent assortment of characters and explained it as follows.

Law of Independent Assortment

"When a dihybrid (or a polyhybrid) forms gametes, (i) each gamete receives one allele from each allelic pair and (ii) the assortment of the alleles of different traits during the gamete formation is totally independent of their original combinations in the parents In other words, each allele of any one pair is free to combine with any allele from each of the remaining pairs during the formation for the gametes

This is known as the Law of Independent Assortment of characters. It is also referred to as *Mendel's third law of heredity.*

Explanation of the law of independent assortment: The principle of independent assortment was explained by Mendel with the help of a dihybrid cross involving characters of cotyledon color (yellow / round) and seed shape *(round / wrinkled).*

Mendel crossed a true breeding variety of pea having yellow cotyledons (YY) and round seeds (RR) with another true breeding variety having green cotyledons (yy) and wrinkled seeds (rr). The complete result of this cross is shown in the Figure 7.3

Thus, the yellow round parent has the genotype (YYRR) and the green wrinkled parent (yyrr). Since each parent is homozygous for both characters (color and shape), each will produce only one type of gametes. The (YYRR) parent will produce all (YR) type gametes and the (yyrr) will produce all (yr) type gametes. All F_1 dihybrids resulting from the fusion of these gametes would be double heterozygous with (YyRr) genotype and appear yellow round. This indicated that in the dihybrid cross also in each pair, the alleles behaved exactly in the same way as in the monohybrid cross. Both the dominants (Y and R) expressed themselves in F_1 while both the recessive alleles (y and r) remained hidden.

Types of gametes formed by F_1 dihybrid: According to Mendel, during gamete formation by the F_1 dihybrid, the alleles in both pairs Y-y and R-r first segregate from each other (Law of segregation). Each pair segregates independently of the pair. Then the alleles enter the gametes. A gamete can receive only one allele from each pair, i.e. Y or y and R or r. Similarly, a gamete that receives a factor (gene) for color must also receive factor for shape (a factor for every character must be present in each gamete). Thus, a gamete that receives Y for color may receive R or r for shape. This would result in (YR) and (Yr) types of gametes. Similarly, a gamete that receives y for color may receive R or r for shape. This would give (yR) and (yr) types of gametes. In other words, the F_1 dihybrid would produce four types of gametes (YR), (Yr), (yR) and (yr) in equal proportions. This is the principle of independent assortment of characters. There will be four types of male gametes and four types of female gametes formed by the F_1 dihybrid.

Test Cross or Back Cross

Test cross (Figure 6) is a simple method devised by Mendel to verify the genotype of the F_1 hybrid. When the F_1 hybrid is crossed with the homozygous recessive parent, it is called a test cross. Since, the F_1 is crossed back with one of the parents, it is also called a back cross.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Test cross is also used for checking the correctness of Mendel's law of segregation (using a monohybrid test cross) and the law of independent assortment of characters (using a dihybrid test cross). For example,

Monohybrid test cross: In a monohybrid cross of *Tall x Dwarf*, the F_1 are all *tall* (Figure 1.1). Let us see what happens when this F_1*tall* is test crossed with the homozygous recessive parent i.e. *Dwarf* with (tt) genotype.

We aim to check and verify two things (i) determine the genotype of F_1 *Tall* and (ii) check the correctness of the law of segregation.

Thus, F_1 *tall* is back crossed (test crossed) with the homozygous recessive *dwarf* parent and the progeny of the cross examined. We know that the recessive dwarf with (tt) genotype will produce only one type of gametes (i.e., all with (t) only). However, as regards the F_1 tall, there can be two possibilities:

If the F_1 *tall* is homozygous with (TT) genotype, it shall produce only one type of gametes (i.e. all with (T) only). As a result, the progeny of the cross should be all tall. (T) x (t) = (Tt) *Tall*.

If, on the other hand, the F_1 tall is heterozygous with (Tt) genotype and, if Mendel's law of segregation of characters is correct, then the F_1 should produce two types of gametes, (T) and (t) in equal proportion. The recessive dwarf parent produces only (T) type of gametes. The resulting test cross progeny should be 50% Tall and 50% Dwarf or in 1:1 ratio.

The results obtained in the actual test cross experiments completely agree with the theoretical expectations. Thus it is proved that F_1 Tall is a heterozygous dominant (monohybrid) with (Tt) genotype and that the alleles segregate during gamete formation.

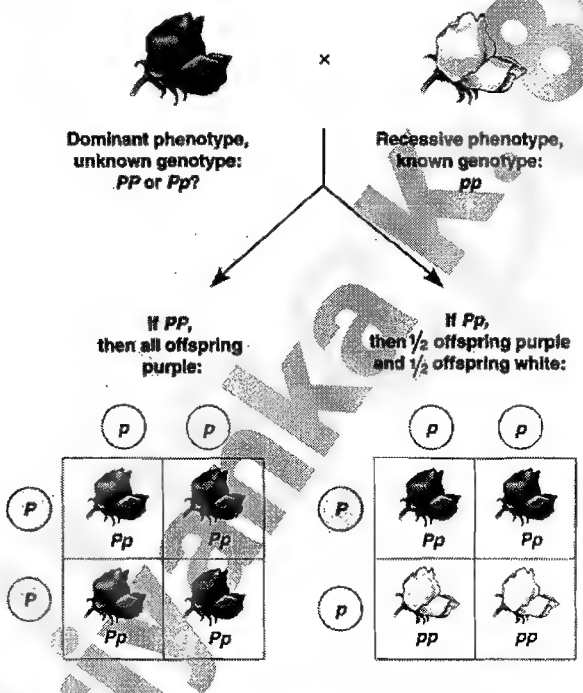


Figure 6: The Concept of Test Cross

Dihybrid test cross: In the dihybrid cross of yellow round X green wrinkled the F_1, dihybrid is a double heterozygous yellow round with (YyRr) genotype. It is test crossed with the double homozygous recessive parent,(i.e. green round with (yyrr) genotype). We aim to test the correctness of Mendel's law of independent assortment of characters. If the principle of independent assortment is correct, then as Mendel explained, the (YyRr) dihybrid should produce four types of gametes: YR, Yr, yR and yr in equal proportions. The recessive parent (yyrr) shall produce only one type of gametes i.e. yr type. It is expected that the maximum possible chance combinations between these gametes should produce four kinds of phenotypes in the ratio of 1:1:1:1. The actual numbers of these four types of offspring obtained in one of Mendel's test crosses were: Yellow round = 55, Yellow Wrinkled = 49, Green Round = 51 and Green Wrinkled = 52. This gives an

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

approximate 1:1:1:1 ratio as per the theoretical expectations and thus, confirms the law of independent assortment of characters.

Significance of back crosses in plant breeding

1. It is a rapid method of improving a variety of crop.
2. It is an easier and quicker method of obtaining a desirable trait in a pure homozygous condition (true breeding trait).
3. Back cross is also use frequently in hybridization experiments because of its simple ratio. Moreover, this needs the analysis of fewer progeny.

Reasons for Mendel's success

1. Mendel concentrated on the results of one trait at a time.
2. When the behavior of one trait was established, only then he considered two characters together.
3. He conducted a large number of crosses (2000-3000) to eliminate the chance factor and to obtain a valid and accurate explanation.
4. Most of all, he actually counted the number of offsprings of each category and maintained accurate records for each generation in each experiment.
5. However, luck played a major role in his success (though Mendel did not know this fact) in the selection of pea plants as well as in the selection of those particular seven pairs of contrasting characters. Because, luckily for Mendel, in each pair, one form of the character is completely dominant over the other.
6. Moreover, these seven characters apparently did not show linkage.

Summary: Mendel's laws of inheritance

1. Genetics is the study of the principles of heredity and variations.
2. Hybridization experiments using garden pea plants by Mendel, and his conclusions and explanations regarding the nature of inheritance of each character are commonly known as Mendelism.
3. Mendel established the phenomena of dominance and recessiveness through monohybrid experiments.
4. Mendel formulated the law of segregation of characters on the basis of the results of monohybrid crosses.
5. The law of independent assortment of characters was formulated by Mendel on the basis of the results of dihybrid crosses.
6. The monohybrid cross in Mendelian experiments gives a phenotypic ratio of 3 : 1 and genotypic ratio of 1:2:1 in the F2 generation.
7. The dihybrid cross gives the phenotypic ratio of 9 : 3 : 3 :1 and genotypic and genotypic dihybrid ratios are the products of their respective monohybrid ratios.
8. Mendel devised the test cross (back cross) method to verify the genotype of F1 hybrid as well as for testing correctness of his laws of heredity.
9. Mendel established the concept of factor, which is now known as gene.

Priyanka 8887661

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 3: Genes, Alleles & Pseusoalleles

Gene

A gene is a hereditary factor and the fundamental unit of inheritance. On the molecular basis, a gene is a piece of DNA that controls a discrete hereditary character. It corresponds usually to a single protein or RNA. The term gene includes the entire functional unit: coding DNA sequences, noncoding regulatory DNA sequences and introns. It is characterized by the existence of different alleles.

Gene concept

From Mendel's experiment on hybridization it is understood that genes are hereditary units transmitted from one generation to other generation through the germ cells and associated with carrying the characters. Mendel used the term **factor** for these hereditary units.

The term gene was first used by Johannsen 1902. After him several workers experimentally proved that genes are the fractions or part of DNA molecule which regarded as the genetic material. Sutton and Boveri independently suggested that chromosome is the container of hereditary units. T.H. Morgan proposed the gene theory which state that:

1. Chromosomes are bearers of hereditary units and each chromosome carries hundreds or thousands of genes.
2. The genes are arranged on the chromosomes in the linear order and on the special regions or locus.

Classical concept of gene

Classical concept of gene was introduced by Sutton (1902) and was elaborated by Morgan (1913). Bridge (1923), Muller (1927) and others, which is outlined as follows.

1. Genes are descrete particles inherited in Mendelian fashion that occupies a definite locus in the chromosome and responsible for expression of specific phenotypic character.
2. Number of genes in each organism is more than the number of chromosomes; hence several genes are located on each chromosome.
3. The genes are arranged in a single linear order like beads on a string.
4. Each gene occupies specific position called locus.
5. If the position or structure of gene changes, character changes.
6. Genes can be transmitted from parent to offsprings.
7. Genes may exist in several alternate formed called alleles.
8. Genes are capable of combined together or can be replicated once during a cell division.
9. Genes may under for sudden changes in position and composition called mutation.
10. Genes are capable of self duplication producing their own exact copies.

Modern concept of gene

1. A gene is a molecular unit of heredity of a living organism.
2. A gene is a specific stretch of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) that code for a polypeptide or for an effector RNA chain.
3. A modern working definition of a gene is "a locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions, and or other functional sequence regions".
4. A gene may be present in a single or multiple copies in the genome.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

5. A gene may have its discrete boundaries or it may be found in an overlapping fashion or in a nested manner.
6. Mostly, genes have numerous alternative forms. They are known as alleles.
7. Through processes such as alternate splicing, exon shuffling and RNA editing, a single gene may give rise to multiple products.
8. Genes are capable of combined together or can be replicated once during a cell division.
9. Genes may under for sudden changes in position and composition called mutation.
10. Genes inherit from a generation to the next through chromosomal replication.

Allele

An allele is one member of a pair of homologous genes in a diploid cell. An individual with identical alleles at a genetic locus is a homozygote; one with non-identical alleles is a heterozygote. In a case in which one allele leads to an observable gene product and the other has no phenotype, the functional allele is said to be dominant and the non-functional allele recessive.

Salient features of alleles are as follows.

1. An allele is one of a number of alternative forms of the same gene.
2. Alleles occupy the same genetic locus.
3. Different alternative forms of a gene for a character produce different effects.
4. Mostly, different alleles result in different observable phenotypic traits. However, many genetic variations result in little or no observable variation.
5. Most multicellular organisms have two sets of chromosomes, that is, they are diploid. These chromosomes are referred to as homologous chromosomes. Diploid organisms have one copy of each gene (and therefore one allele) on each chromosome. If both alleles are the same, they and the organism are homozygous, and the organisms are homozygotes. If the alleles are different, they and the organism are heterozygous and the organisms are heterozygotes.
6. A population or species of organisms usually have multiple alleles at each locus among various individuals.
7. Different alleles for the same gene mostly arise from intra-genic point mutations.

Pseudoallele

Pseudoalleles refer to the genes, which behave as alleles in allelism test but can be separated by crossing over. Thus, a pseudo-allele is a different gene that is seemingly allelic to another gene. However, this gene can be shown to have distinctive locus. Such a locus is often closely linked.

True alleles, in contrast to the pseudoalleles, refer to different versions of one gene only, and they always occupy the same locus on the chromosome. True allelism thus relates to one of a series of two or more alternate forms of a gene that occupy the same position or locus on a specific chromosome.

Salient features of the pseudoalleles would include the following.

1. They are two different genes, rather than two different versions of the same gene. Nevertheless, they are related to the same phenotypic character of the individual. Therefore, pseudoalleles can be called functionally but not locationally allelic.
2. They occupy two different loci on the chromosome.
3. Pseudoalleles are mostly located on the same chromosome; hence, they fail to give the segregation ratio as predicted by the Mendel's Law of Independent Assortment. The pseudoalleles are not separated during meiosis, unless a crossing-over takes place.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

4. Pseudoalleles mostly arise from gene duplications. Gene duplication creates a gene copy immediately next to the original gene. For this reason, generally, the pseudoalleles are tightly linked to each other on the same chromosome.
5. An allele and its respective pseudoallele may accumulate mutations independently and thus go through autonomous evolutionary processes.
6. The effects of an allele and its pseudoallele are only rarely additive. A pseudoallele does thus not replace an allele.

The origin of pseudoalleles is mostly thought to have taken place by ***gene duplication*** or ***illegitimate crossing overs***. Gene duplication creates a gene copy immediately next to the original gene.

Illegitimate crossing-overs result in two products of unequal length. Consequently, one chromatid loses a fragment (a deletion) to its partner chromatid. This does now harbour the fragment twice (a duplication). Usually, duplicated fragments are arranged in a tandem. The classic example for this simultaneous deletion/duplication event is seen in the the *Bar*-gene of *Drosophila* (the *Bar*-locus; *Bar* is written with a capital B due to its dominance over the wild type). The eyes of *Drosophila* are normally roughly oval. However, those of the *Bar*-mutant are shaped like a bar. This phenotype is caused by a duplication. The expression of the characteristic *Bar* shows that the surplus of an additional locus can also cause phenotype changes apart from the lack or the defect of a gene.

These pseudoalleles are not separated during meiosis, unless a crossing-over takes place. The pseudoalleles, being related to the same function and located on the same chromosome, affect the gene function in a significant way, some times. This and other observations lead to the discovery that the genes of a genome are no independent units. Their activities are actually controlled by neighbouring DNA domains. This phenomenon was originally called the position effect but quite often it may be due to pseudoalleles.

Priyanka. 8023871651

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 4: Multiple allelism

Alleles

Allele constitutes one of two or more alternative forms of a gene at a given position (locus) on a chromosome, caused by a difference in the sequence of DNA. This is best explained with examples. A gene, which controls eye colour in humans, may have two alternative forms - an allele that can produce blue eyes, and an allele that produces brown eyes. In a plant that occurs in tall and short forms, there may be an allele that tends to produce tall plants and an alternative allele that produces short plants.

The individual genes that form a pair of alleles are located at exactly the same point along a chromosome. Organisms with two sets of chromosomes (diploids), such as animals and plants, have chromosomes that are found as matching pairs in the nucleus of each cell. This means that there will always be two genes for a characteristic in a cell. If the same allele is present twice, the organism is said to be homozygous for this characteristic. If, however, one chromosome contains one allele and the other chromosome a contrasting allele, the organism is said to be heterozygous.

In a heterozygous organism the appearance of the organism (phenotype) may be determined by one allele and not the other. The allele that determines the phenotype is said to be dominantly expressed; it shows dominance over other alleles. The expression of the other allele is described as being recessive.

A general account multiple allelism

A gene for which at least two alleles exist is said to be **polymorphic**. Instances in which a particular gene may exist in three or more allelic forms are known as **multiple allele conditions**. It is important to note that while multiple alleles occur and are maintained within a population, any individual possesses only two such alleles (at equivalent **loci** on **homologous** chromosomes).Among humans, certain traits, such as blood types, hair color, and eye color, have more than two alleles that are found in a population.

Multiple allelism is fundamentally different from polygenic inheritance, which can be explained as follows.

Alleles are different versions or forms of ONE gene. For example, there can are 3 alleles (A, B, and O) that contribute to blood types in humans. Each person can have only two versions of the trait in their genome at one time, but in the population, there are three forms. In fruit flies, there are many different versions of eye color in the population, but one fly can have only 2 alleles at a time.

Polygenic traits are those where more than one GENE contributes to the phenotype. For example, the current model of skin color is that there are three genes that contribute to skin color. So there are 6 total alleles, 3 from the mother and 3 from the father.

Origin of multiple alleles

Multiple alleles can develop due to the following three reasons.

1. **Mutation of Gene**. Because the information stored in any gene is extensive, mutations can modify the gene in many ways. Each change has the potential for producing a different allele. Therefore, for any gene, the number of alleles within members of a population of individuals is not necessarily restricted to two. When three or more alleles of the

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

same gene are found, multiple alleles are said to be present, creating a characteristic mode of inheritance.

2. **Pseudo alleles.** Multiple alleles represent components of a complex locus having pseudo alleles.
3. **Transposition of Heterochromatin.** Minute fragments of heterochromatin may shift or get deposited near the genes bringing about change in their expression.

Characteristics

1. All the multiple alleles occur at the same gene locus in the same chromosome type.
2. A chromosome contains only single allele.
3. An individual contains only two alleles of multiple allele series. They occur on the two chromosomes of a homologous pair. The two alleles segregate at the time of gamete formation and come together at the time of fertilization.
4. Multiple alleles of a gene affect the same character.
5. They affect the different alternatives of a character.
6. Out of the different multiple alleles, usually one is normal or of wild type. All others are its mutants.
7. The normal or wild type allele is commonly dominant over all other alleles.
8. Amongst themselves, the mutant alleles show a relationship of dominance-recessiveness or intermediate dominance. Consequently, an individual having two different mutant alleles (compound or heterozygote) may have one mutant phenotype or phenotype intermediate between the expressions of the two mutant alleles. It is never of wild type.

Examples of multiple allelism

1. Three well studied human examples of multiple-allele genes are the gene of the ABO blood group system, the human-leukocyte-associated antigen (HLA) genes and Polymorphism in Noncoding Dna.

Genotype	Blood Group
$I^A I^A$	A
$I^A I^O$	A
$I^B I^B$	B
$I^B I^O$	B
$I^A I^B$	AB
$I^O I^O$	O

   a. The ABO system in humans is controlled by three alleles, usually referred to as I^A, I^B, and I^O (the "I" stands for isohaemagglutinin). I^A and I^B are codominant and produce type A and type B antigens, respectively, which migrate to the surface of red blood cells, while I^O is the recessive allele and produces no antigen. The blood groups arising from the different possible genotypes are summarized in the adjoining table.

   b. HLA genes code for protein antigens, which are expressed in most human cell types and play an important role in immune responses. These antigens are also the main class of molecule responsible for organ rejections following transplantations—thus their alternative name: major histocompatibility complex (MHC) genes. The most striking feature of HLA genes is their high degree of polymorphism—there may be as many as one hundred different alleles at a single locus. If we consider that an individual possesses five or more HLA loci, it becomes clear why donor-recipient matches for organ transplantations are so rare (the fewer HLA antigens the donor and recipient have in common, the greater the chance of rejection).

   c. Polymorphism in Noncoding Dna: From the results of the Human Genome Project, it is well established that Most of the DNA sequence variation between individuals

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

arises not because of differences in the genes, but because of differences in the noncoding DNA found between genes. An example of a noncoding DNA sequence that is extremely abundant in humans is the so-called microsatellite DNA. Microsatellite sequences consist of a small number of nucleotides repeated up to twenty or thirty times.For instance, the microsatellite composed of the dinucleotide AC is very common, appearing about one hundred thousand times throughout the human genome. The interesting feature about microsatellites is that they are very highly polymorphic from one individual to another for the number of repeat lengths. For example, one particular individual might possess the microsatellite sequence ACACACACAC at a specific locus on one chromosome, and the other individual may have the sequence ACACACACACACACACAC at the same locus the same chromosome. Noncoding polymorphic DNA is of considerable importance in gene mapping and DNA fingerprinting.

2. Other well-studied examples of multiple allelism would include the following.

   a. In *Drosophila,* many alleles are present at practically every locus. The recessive eye mutation, *white,* discovered by Thomas H. Morgan and Calvin Bridges in 1912, represents only one of over 100 alleles that can occupy this locus. In this allelic series, eye colors range from complete absence of pigment in the *white* allele, to deep ruby in the *white-satsuma* allele, to orange in the *white-apricot* allele, to a buff color in the *white-buff* allele. In each case, the total amount of pigment in these mutant eyes is reduced to less than 20 percent of that found in the brick-red wild-type eye.

   b. The various coat colors found in rabbits is also controlled by multiple alleles. The dominant color gene is the "Full Color" gene represented by the capital "C." Some full color rabbits are black, orange, chestnut, black tortoiseshell, lilac, and blue. The next gene, in order of dominance, is the chinchilla gene, which is represented by the letters *chd* (or sometimes *Cchd* or c_{chd}). The letters stand for "chinchilla-dark." The next gene is the sable gene represented by c_{chl}, which stands for "chinchilla-light." Smoke pearl and sable point are examples of rabbits with the chl as the dominant color gene. Himalayans have the ch (or c_h) gene while ruby-eyed whites (REWs) have two "c" genes.

Importance of multiple allelism

1. Multiple alleles prove that a gene can mutate many times. The mutated gene can mutate further.
2. A nonallelic gene can mutate in such a fashion as to directly influence the trait normally controlled by another.
3. Each gene occurs on a particular locus on its particular chromosome.
4. Multiple alleles provide increased variability even in a single trait.
5. With the help of self-sterility alleles, multiple alleles prevent inbreeding.
6. Multiple alleles increase competitiveness of individuals in struggle for existence and enhance chances of natural selection.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 5: Incomplete dominance

A general account of dominant recessive relationship

In genetics, **dominance recessive relationship** describes the effects of the different versions of a particular gene on the phenotype of an organism. All diploid organisms have two copies of each gene in their genome, one inherited from each parent. The different variants of a specific gene are known as alleles. If an organism inherits two alleles which are different from one another, and the phenotype of the organism is determined completely by one of the alleles, then that allele is said to be **dominant**. The other allele, which has no tangible effect on the organism's phenotype, is said to be **recessive**.

Dominance was discovered by Mendel, who introduced the use of uppercase letters to denote *dominant alleles* and lowercase to denote *recessive alleles*, as is still commonly used.

In the simple example of flower color in peas, first studied by Gregor Mendel, the dominant allele is purple and the recessive allele is white. In a given individual, the two corresponding alleles of the chromosome pair fall into one of three patterns:

- both alleles purple (*PP*) – the phenotype would be purple flowers
- both alleles white (*pp*) – the phenotype would be white flowers
- one allele purple and one allele white (*Pp*) – the phenotype would be purple flowers

In the above example, if the two alleles are the same (homozygous), the trait they represent will be expressed. However, if the individual carries one of each allele (heterozygous), only the dominant one will be expressed. The recessive allele will simply be suppressed.

In most but not all the cases, a dominance relationship is seen when the dominant version of the gene encodes an enzyme, and its recessive counterpart does not. In some cases, the recessive copy makes the enzyme but in lesser amounts or makes a different or abberent version of the enzyme.

Incomplete dominance

Discovered by Carl Correns, Incomplete Dominance (sometimes called Partial Dominance or Intermediate Inheritance) is a situation in which heterozygous genotype creates an intermediate phenotype. In this case, only one allele (usually the wild type) at the single locus is expressed in a doseage dependent manner, which results in an intermediate phenotype. Incomplete or intermediate dominance has traditionally been described as a phenomenon where none of the two Mendellian factors or alleles is dominant over the other.

A cross of two intermediate phenotypes (= monohybrid heterozygotes) will result in the reappearance of both parent phenotypes and the intermediate phenotype. There is a 1:2:1 phenotype ratio instead of the 3:1 phenotype ratio found when one allele is dominant and the other is recessive. This allows an organism's genotype be diagnosed from its phenotype without time-consuming breeding tests.

Incomplete dominance is an example of modifications of Mendellian ratios due to interallelic (that is intragenic) interaction, which can be defined as the influence of one allele over another which causes a change in its expression, phenotype or normal Mendellian ratios 3 : 1 (monohybrid cross) and 9: 3 : 3 : 1 (dihybrid cross).

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Some Well Studied Examples

1. In ***Mirabilis jalapa*** (Four O'clock plant) there are two types of flower colours in the pure state; red (RR) and white (rr). When the two types of plants are crossed, the hybrids or plants of F 1 generation have pink flowers (Fig. 1) showing that neither of the two genes is dominant over the other though the effect of one gene (for red flower) is more pronounced than the other. The Pink colour appears to be formed by mixing of red and white colours. That the pink colour is produced by the joint activity of the two genes is confirmed by selfing the hybrids and producing F_2 generation. In F 2 generation the plants are of three types - red flowered, pink flowered and white flowered in the ratio of 1:2:1. This is a typical genotypic Mendellian ratio, showing that pink colour is not produced by any specific gene but by the joint activity of two alleles.

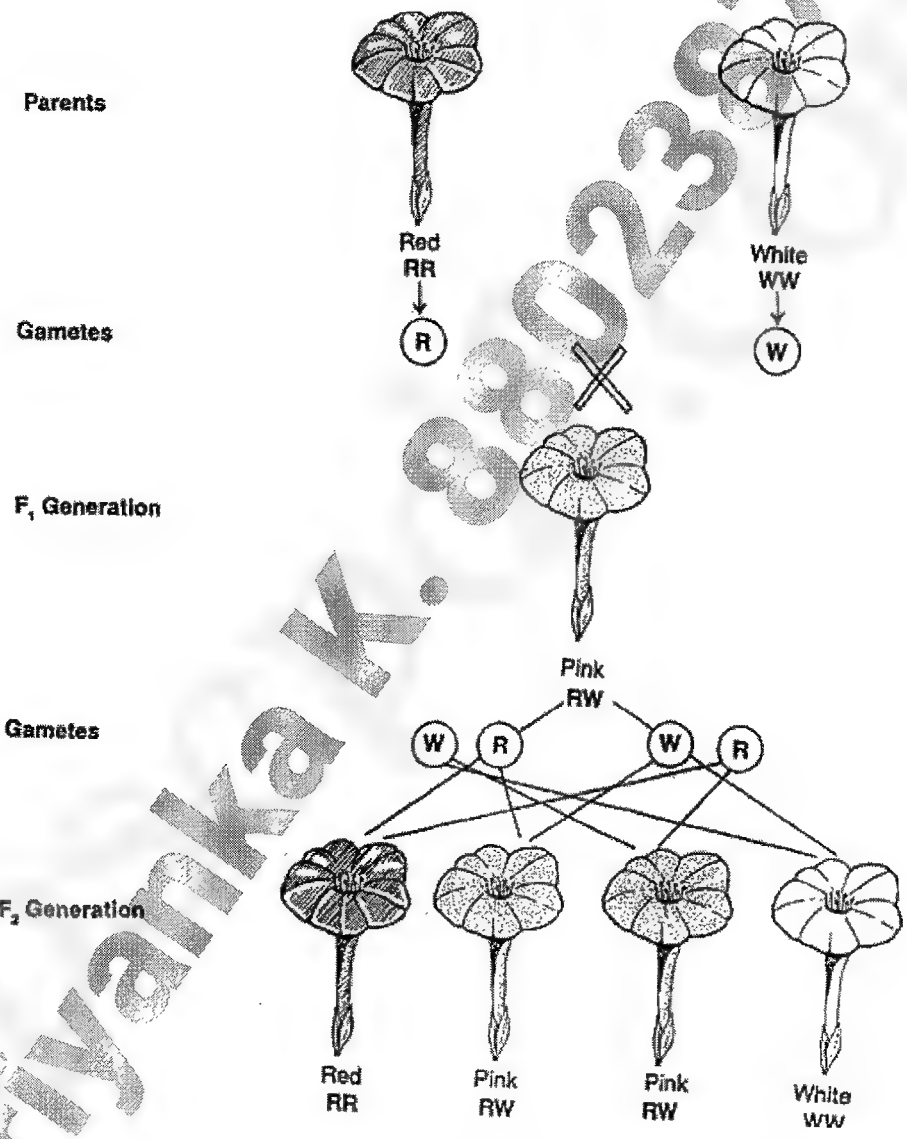


Figure 1 : Inheritance of flower colour in *Mirabilis jalapa* indicating incomplete dominance

2. **Andalusian fowls** have two pure forms, black (BB) and white (bb). If the two types of fowls are crossed, F 1 or hybrids are blue coloured. The blue colour appears due to fine mixing of black and white stripes on the feathers. Because of the presence of both the

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

colours, some zoologists consider it a case of codominance. On inbreeding the blue fowls give rise to three types of fowls black, blue and white in the ratio of 1 : 2 : 1 (Fig.3). Black and white colours do not mix up in the F 1 hybrids, because they segregate in the F 2 fowls. Therefore, blue colour is produced due to activity of alleles of both black and white colour.

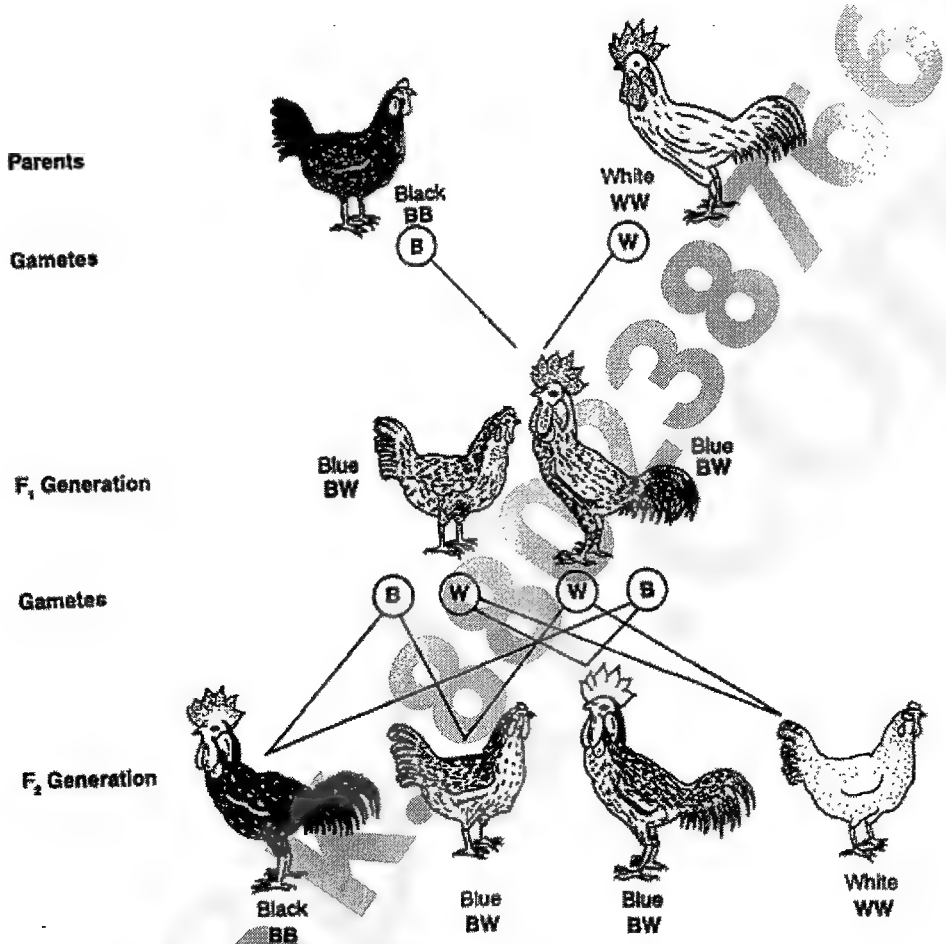


Figure 2: Inheritance of feather colour Andalusian fowls

3. There are two types of pure short horned, polled or hornless cattle, red haired (RR) and grey or white (rr). On being crossed, they produce hybrids having roan colour. The roan colour is not formed by intermixing of red and grey hair. Instead, it is produced by small patches of red and white colour. Because of it, the inheritance of fur colour in hornless cattle is often also considered a case of codominance. On inbreeding roan hybrids produce three types of cattle- red, roan and white in the ratio of 1 : 2 : 1 indicating that neither the traits (red and grey) nor their alleles mix up in the F 1 generation because they reappear in the F_2 individuals.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

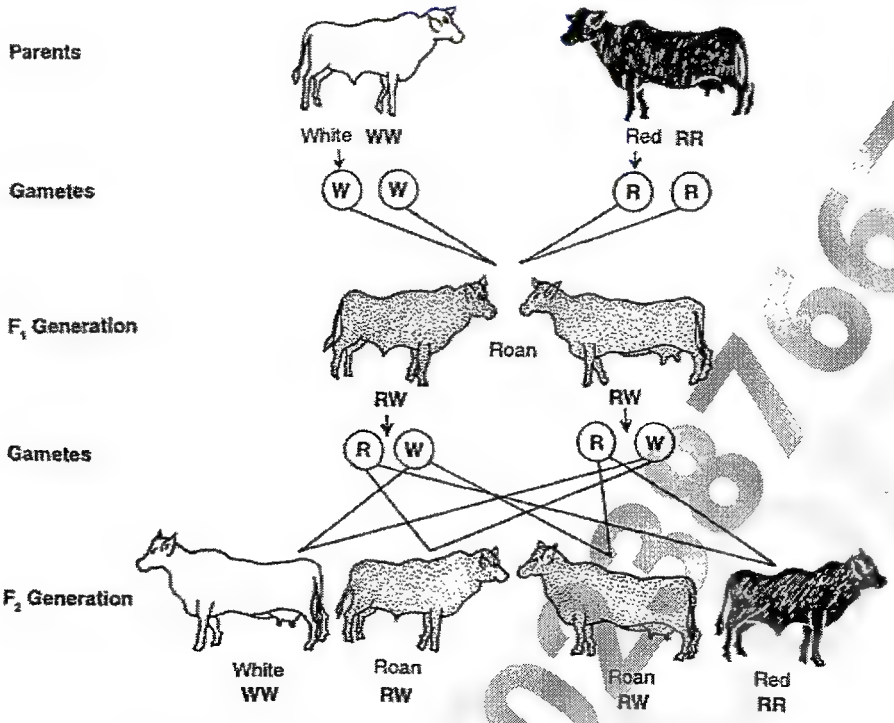


Figure 3: Inheritance of hair colour in polled cattle

4. In *Antirrhinum* (Snapdragon or Dog Flower), there are two traits which show incomplete dominance - flower colour and breadth of leaves. Plants with broad leaves (BB) and narrow leaves (NN) yield intermediate leaves (BN) on crossing. Similarly, a cross between plants having red flowers (RR) and white flowers (WW) yield pink flowered plants (RW). A dihybrid cross (BBRR x NNWW) yields F 1 plants having pink flowers and intermediate leaves. On self breeding, nine types of F 2 plants are obtained in the ratio of 1 : 2 : 1 : 2 : 4 : 2 : 1 : 2 : 1 .

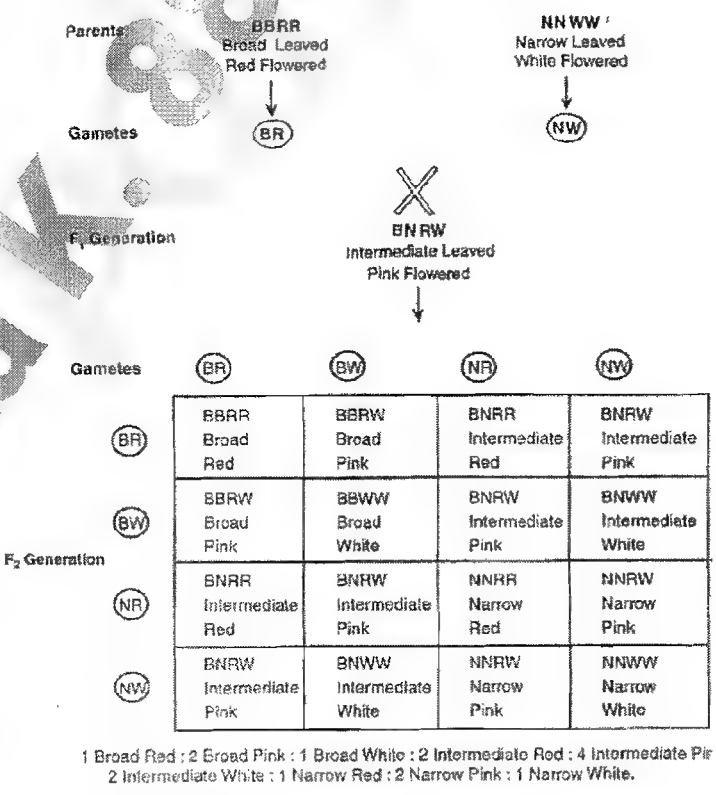


Figure 4: Inheritance of two traits with incomplete dominance in *Antirrhinum*

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 6: Polygenic traits and quantitative inheritance

Introduction to qualitative and quantitative inheritance

Qualitative inheritance (Monogenic Inheritance). It is the type of inheritance in which a single dominant gene influences a complete trait. Presence of two such dominant copies of a gene does not alter the phenotype. The genes controlling inheritance are called **monogenes** *e.g.* TT or Tt for tallness in Pea. Qualitative inheritance produces a sort of *discontinuous trait variations in* the progeny, *e.g.,* either tallness or dwarfness. Intermediate forms or continuous trait variations are not produced.

Qnantitative Inheritance (Polygenic Inheritance). *It is a type of inheritance controlled by more than one genes in which the dominant alleles of different genes have cumulative effect. Here each dominant allele expresses a part or unit of the trait, and the full trait is shown only when all the dominant alleles are present. The genes involved in quantitative inheritance are called* ***polygenes***. Quantitative inheritance is, therefore, also called **polygenic inheritance.** It is also named as **multiple factor inheritance.**

A few instances of quantitative inheritance are:

1. kernel colour in wheat
2. cob length in Maize
3. skin colour in human beings
4. human intelligence
5. milk and meat yield in animals
6. height in human beings and several plants
7. yield of crop plants including size, shape and number of seeds or fruit per plant.

A **polygene** is defined as a gene where a dominant allele controls only a unit or partial quantitative expression of a trait. It is also termed as a gene in which a dominant allele individually produces a slight effect on the phenotype but in the presence of similar other dominant allele it controls the quantitative expression of a trait due to cumulative effect. Hence, polygenes are also called **cumulative** genes. The possible origin of polygenes is (i) Duplication of chromosome part (ii) Polyploidy or increase in chromosome number (iii) Mutations producing genes having similar effect.

The traits controlled by quantitative inheritance are sometimes known as **metric traits** because they can be measured in terms of unit of size, height, weight or number. Quantitative inheritance is further characterized by the occurrence of intermediate forms (**continuous variations**) between the parental types (Fig. 1). Here a cross between two pure breeding parents does not produce dominant trait of one parent but instead an intermediate trait is exhibited. Similarly, in F_2 generation apart from the two parental types there are several intermediate types, which link the two parental traits. Because of the latter, quantitative inheritance is also called **blending inheritance.**

The dominant polygenic alleles which contribute to the expression of the trait are called **contributing alleles** while the recessive polygenic alleles are known as **non-contributing alleles.**

In natural populations, variation in most characters takes the form of a continuous phenotypic range rather than discrete phenotypic classes. In other words, the variation is quantitative, not qualitative. Mendellian genetic analysis is extremely difficult to apply to such continuous

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

phenotypic distributions, so statistical techniques are employed instead – an approach known as **Quantitative Genetics.**

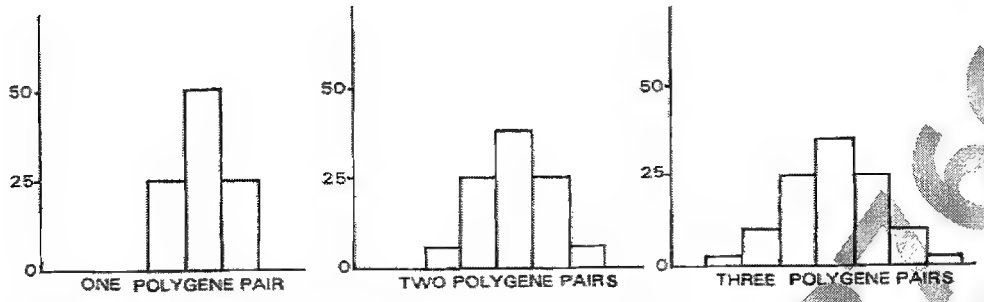


Figure 1 : Histograms showing the distribution of F2 phenotypes in case of polygenic inheritance.

Early studies on quantitative inheritance

Quantitative or polygenic inheritance was first studied by **J. Kolreuter** (1760) in case of tobacco and **F. Galton** (1883) in case of human beings. **Nilsson-Ehle** (1908) obtained the first experimental proof of polygenic inheritance in case of kernel colour in wheat.

Salient features of quantitative inheritance

1. Many—perhaps most—of the phenotypic traits that we observe in organisms vary continuously. In many cases, the variation of the trait is determined by *more than a single* segregating locus.
2. Each of these loci may contribute equally to a particular phenotype, but it is more likely that they contribute unequally. The measurement of these phenotypes and the determination of the contributions of specific alleles to the distribution is made on a statistical basis in these cases.
3. Some of these variations of phenotype (such as height in some plants) may show a *normal distribution* around a mean value; others (such as seed weight in some plants) will illustrate a *skewed distribution* around a mean value.
4. In other characters, the variation in one phenotype may be correlated with the variation in another. A correlation coefficient may be calculated for these two variables.
5. With the use of genetically marked chromosomes, it is possible to determine the relative contributions of different chromosomes to variation in a quantitative trait, to observe dominance and epistasis from whole chromosomes, and, in some cases, to map genes that are segregating for a trait.
6. Traits are called familial if they are common to members of the same family, for whatever reason.
7. Traits are called heritable, only if the similarity arises from common genotypes. In experimental organisms, environmental similarities may be readily distinguished from genetic similarities, or heritability.

Detection of polygenic inheritance

Polygenic inheritance can be known from the frequency distribution of phenotypes. In monogenic or qualitative inheritance the phenotypes are two (3: 1) in case of single gene pair and 4 (9: 3: 3: 1) in case of two pairs of genes.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

In polygenic or quantitative inheritance the number of phenotypes is 3 (1 : 2 : 1) in case of one polygene pair, 5 (1 : 4 : 6 : 4 : 1) in case of two polygene pairs and 7 (1 : 6 : 15 : 20 : 15 : 6 : 1) when three polygene pairs are involved. Thus, we see that the number of intermediate types increases with the increase in the number of polygenes but the number of parental types remain the same (two in the above cases).

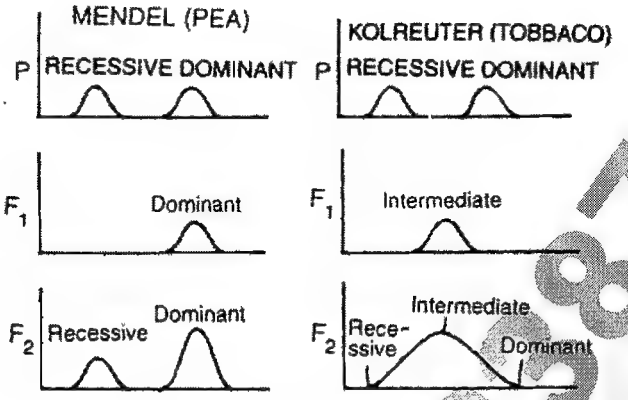


Figure 2: Difference between monogenic (qualitative) and polygenic (quantitative) in heritance

Differences between monogenes and polygens

Monogenes/Monogenic inheritance	*Polygenes/Polygenic inheritance*
1. They produce discontinuous variations in the expression of traits.	1. Polygenes produce continous variations in the expression of traits.
2. A single dominant allele expresses the complete trait.	2. A single dominant allele expresses only a unit of the trait.
3. Monogenic inheritance controls qualitative traits.	3. Polygenic inheritance controls quantitative or metric trait.
4. A character is represented in an individual by a pair of alleles.	4. A character is represented by one to several pairs of alleles.
5. F_1 individuals are similar to dominant parent.	5. F_1 individuals are intermediate between the two parents.
6. F_2 individuals resemble both the parents in the ratio of 3 : 1.	6. Depending upon the number of polygenes, 2/4 (one pair), 2/16 (two pairs) or 2/64 (three pairs) F_2 individuals resemble the parental types.
7. No intermediates are produced in monogenic or qualitative inheritance.	7. Intermediates are quite common in polygenic or quantitative inheritance.
8. There is no cumulative action in the presence of two dominant genes.	8. The dominant genes have cumulative effect on the expression of the trait.
9. Individuals with dominant phenotype are more numerous than with recessive phenotype.	9. Individuals with dominant trait are usually as few as with recessive trait. Intermediate forms are more numerous.

The principles of quantitative genetics

The phenotypic value (P) of an individual is the combined effect of the genotypic value (G) and the environmental deviation (E):

$$P = G + E$$

The genotypic value is the combined effect of all the genetic effects, including nuclear genes, mitochondrial genes and interactions between the genes. It is therefore often subdivided in an additive (A) and a dominance component (D). The additive effect described the cumulative effect

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

of the individual genes, while the dominance effect is the result of interactions between those genes. The environmental deviation can be subdivided in a pure environmental component (E) and an interaction factor (I) describing the interaction between genes and the environment. This can be described as:

$P = A + D + E + I$

The contribution of those components cannot be determined in a single individual, but they can be estimated for whole populations by estimating the variances for those components

Some well known examples of quantitative inheritance

1. **Kernel color in wheat** is determined by two gene pairs, so called polygenes that produce a range of colors from white to dark red depending on the combinations of alleles. Dark red plants are homozygous: AABB and white plants are homozygous aabb. When these homozygotes are crossed the F1 offspring are all double heterozygotes AaBb. Thus crossing individuals with the phenotype extremes yield offspring that are a blend of the two parents.

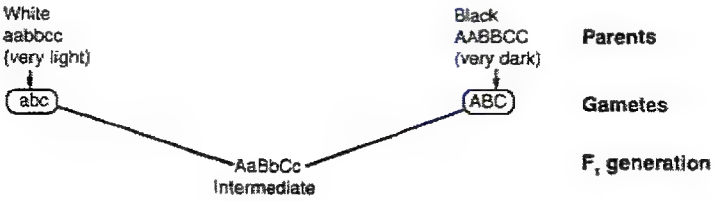
   The results obtained when the two double heterozygotes are crossed are shown in the following Punnett Square

	AB	**Ab**	**aB**	**ab**
AB	AABB (dark red)	AABb (red)	AaBB (red)	AaBb (Faint red)
Ab	AABb (red)	AAbb (faint red)	AaBb (faint red)	Aabb (pink)
aB	AaBB (red)	AaBb (faint red)	aaBB (faint red)	aaBb (pink)
ab	AaBb (faint red)	Aabb (pink)	aaBb (pink)	aabb (white)

2. **Human Skin Colour**. It was first studied by Davenport (1913) in case of Negro-caucasian intermarriages in Jamaica and Muda. Human skin colour is caused by pigment called melanin. The quantity of melanin is due to three pairs of polygenes (A, B and C). If black or very dark (AABBCC) and white or very light (aabbcc) individuals marry, the offspring or individuals of Fl generation show intermediate colour often called mulatto (AaBbCc). When two such individuals of intermediate colour marry, the skin colour of the children will vary from very dark or black to very light or white. A total of eight allele combinations is possible in the gametes forming 27 distinct genotypes distributed into 7 phenotypes. (Fig 3)

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf



Gametes → / ↓	ABC	aBC	AbC	ABc	abC	Abc	aBc	abc
ABC	AABBCC very dark	AaBBCC dark	AABbCC dark	AABBCc dark	AaBbCC fairly dark	AABbCc fairly dark	AaBBCc fairly dark	AaBbCc intermediate
aBC	AaBBCC dark	aaBBCC fairly dark	AaBbCC fairly dark	AaBBCc fairly dark	aaBbCC intermediate	AaBbCc intermediate	aaBBCc intermediate	aaBbCc fairly light
AbC	AABbCC dark	AaBbCC fairly dark	AAbbCC fairly dark	AABbCc fairly dark	AabbCC intermediate	AAbbCc intermediate	AaBbCc intermediate	AabbCc fairly light
ABc	AaBBCc dark	AaBBCc fairly dark	AABbCc fairly dark	AABBcc fairly dark	AaBbCc intermediate	AABbcc intermediate	AaBBcc intermediate	AaBbcc fairly light
abC	AaBbCC fairly dark	aaBbCC intermediate	AabbCC intermediate	AaBbCc intermediate	aabbCC fairly light	AabbCc fairly light	aaBbCc fairly light	aabbCc light
Abc	AABbCc fairly dark	AaBbCc intermediate	AAbbCc intermediate	AABbcc intermediate	AabbCc fairly light	AAbbcc fairly light	AaBbcc fairly light	Aabbcc light
aBc	AaBBCc fairly dark	aaBBCc intermediate	AaBbCc intermediate	AaBBcc intermediate	aaBbCc fairly light	AaBbcc fairly light	aaBBcc fairly light	aaBbcc light
abc	AaBbCc intermediate	AaBbCc fairly light	AabbCc fairly light	AaBbcc fairly light	aabbCc light	Aabbcc light	aaBbcc light	aabbcc very light

Figure 3 : Quantitative inheritance of skin colour in human beings

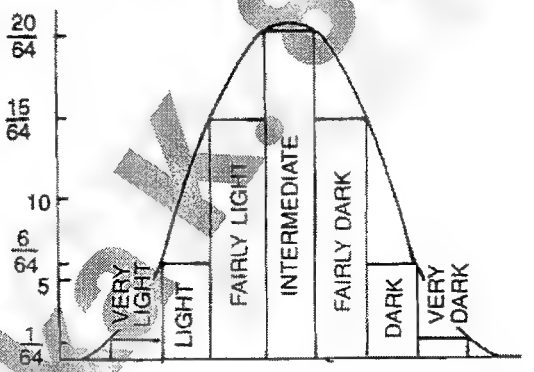


Figure 4 : A histogram prepared from the frequencies of various phenotypes showing a bell shape curve

Priyanka's Study Point

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 7: Linkage & Linkage Mapping

Introduction

In any organism, the total number of genes is always much more than the number of chromosomes, the physical carriers of the genes. It is obvious that every single chromosome must carry a large number of genes. An average human chromosome, for instance, has more than one thousand genes on it.

Genetic linkage is a phenomenon where *two or more genes tend to be inherited together in a fixed allelic combination from one generation to the next, due to the fact that they are located on the same chromosome.* All the genes located on the same chromosome are said to be **linked genes** and they together make one **linkage group**. The maximum number of linkage groups in an organism is equal to its unique chromosome number (that is haploid autosome number + the types of heterosomal sex chromosomes).

Genetic linkage *causes deviations from the expected Mendellian dihybrid ratios* because linked genes often fail to show independent assortment, a necessary precondition for obtaining the 9:3:3:1 dihybrid ratio. Microtubules are macromolecular assemblies of globular proteins.

Discovery and Characterization

The phenomenon was first observed in 1905 by **William Bateson** and **R. C. Punnett** when they were studying inheritance in the plant *Lathyrus odoratus* (sweet pea). They studied the transmission of two genes:

One affecting flower color (*P*, purple, and *p*, red) and the other affecting the shape of pollen grains (*L*, long, and *l*, round).

They crossed pure lines *P*/*P* · *L*/*L* (purple, long) × *p*/*p* · *l*/*l* (red, round), and selfed the F_1 *P*/*p* · *L*/*l* heterozygotes to obtain an F_2.

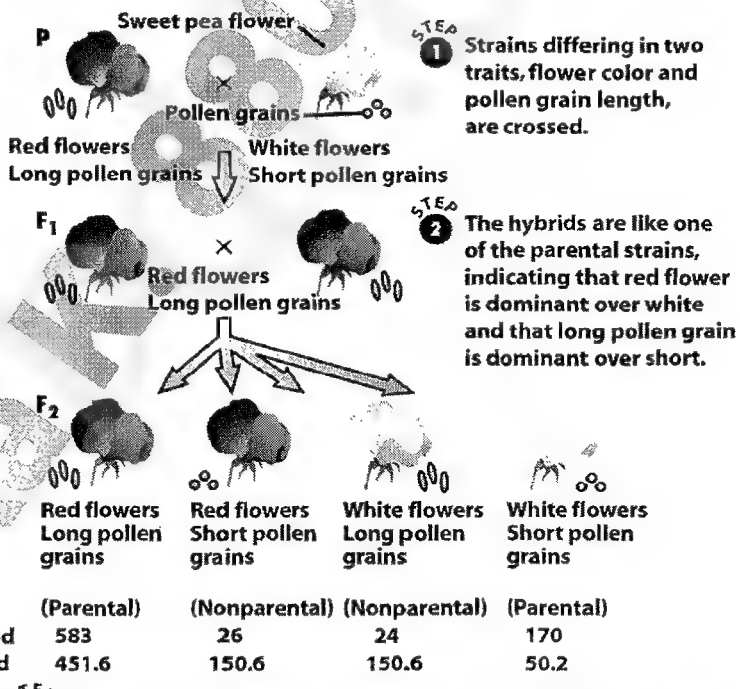


Figure 7: The Cross in *Lathyrus odoratus* that led to the discovery of Linkage

The F_2 phenotypes deviated strikingly from the expected 9:3:3:1 ratio (**Fig.** 1) There were far more parental types (*P*/*P* . *L*/*L* and *p*/*p* . *l*/*l*) and far less recombinant types (*p*/*p* . *L*/- or *P*/- .

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

l/l). The data also did not appear to be explainable as a modified Mendelian ratio or any other principle of genetic transmission known at that time.

This was an outcome of linkage, that is the two genes in question were linked – hence failed to show independent assortment. However, *Bateson & Punnet provided an inaccurate explanation* for what they had observed. The proposed two theories to explain this mode of genetic transmission, both of which were proved to be wrong by subsequent findings. The two theories were:

Coupling and Repulsion Hypothesis: According to this, there exist two types of relation between two genes.

Coupling: In which the dominant allele of one gene would tend to be inherited with the dominant allele of another gene. Similarly, the recessive allele of the first gene would tend to be inherited with the recessive allele of second gene. Thus, F_1 produces more $P \cdot L$ and $p \cdot l$ gametes than what would be produced by Mendellian independent assortment.

Repulsion: In this situation, the nonallelic dominant alleles "repelled" each other and also the the nonallelic recessive alleles "repelled" each other. This relation is exactly opposite to the coupling relation.

Preferential gamete formation: According to this model, the chromosomes fragment at the time of gamete formation and when they re-assemble there is a greater tendency of parental allelic combinations being retained in the gametes too.

It must be noted that both the above hypotheses have been proven wrong. *The correct explanation of the phenomenon was provided by* **T.H. Morgan** *and co-workers in 1910s.*

In the course of their research on *Drosophila* genetics, Morgan and his students also observed that some pairs of genes did not segregate randomly according to Mendel's principle of independent assortment but instead tended to be inherited together. Morgan suggested that the genes **are located on the same chromosome and thus traveled together during** meiosis. This is the correct explanation to the phenomenon.

Later, Morgan and his students also discovered the phenomenon of meiotic crossing over and developed a method of gene mapping based on the frequency of crossing over between two linked genes.

Salient Features

The main features of linkage are given below:

1. Linkage involves two or more genes which are located in the same chromosome in a linear fashion.
2. Linkage may involve either dominant genes or recessive genes or some dominant and some recessive genes.
3. Linked genes definitely deviate from independent assortment but not in 100% cases.
4. Independent assortment between linked genes occurs by crossing over. If crossing over does not occur, all the genes located in one chromosome are expected to be inherited together.
5. The stability of linkage is measured in terms of how much the linked genes fail to show independent assortment. If independent assortment (by crossing over) frequency is low, the linkage is said to be stable or stronger.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

6. The strength of linkage depends on the distance between the linked genes. Lesser the distance higher the strength and vice versa. So, stable linkage usually involves those genes which are located closely.
7. Presence of linkage leads to higher frequency of parental types than recombinants in a test cross progeny. When two genes are linked the segregation ratio of a test cross progeny deviates significantly from the 1: 1: 1: 1 ratio.
8. Linkage is an "all-applicable" phenomenon; hence it may involve either two or more desirable traits or all undesirable traits or some desirable and some undesirable traits.
9. Besides pleiotropy, linkage is an important cause of genetic correlation between various genetic characters.
10. Linkage can be broken by repeated intermating of randomly selected individuals in segregating populations for several generations.

Types of Linkage

Linkage is generally classified on the basis of three criteria, *viz.*, (1) presence or absence of crossing over, (2) genes involved, and (3) the chromosome involved.

1. **Based on crossing over**

   - **Complete Linkage**. Linkage in which crossing over does not occur is known as complete linkage or absolute linkage. In other words, when only parental types are obtained from the test cross progeny, it refers to complete linkage. Usually, no linkage is considered to be absolutely complete.
   - **Incomplete Linkage**. If some frequency of crossing over also occurs between linked genes, it is known as incomplete linkage. To put in other way, when recombinations are also observed in the test cross progeny, besides parental combinations, it refers to incomplete linkage. Incomplete linkage is the most common type of linkage.

2. **Based on Genes Involved**

   - **Coupling Linkage**. It refers to linkage either between dominant genes or between recessive genes.
   - **Repulsion Linkage**. It refers to linkage of one dominant gene with another recessive gene. This type of linkage has also been observed in pea, maize and several other crops.

3. **Based on Chromosome Involved**

   - **Autosomal Linkage**. It refers to linkage of such genes which are located in other than sex chromosomes (autosomes).
   - **Sex Linkage**. It refers to the linkage of genes which are located in sex chromosomes.

Detection of Linkage

There are **two simple methods** of linkage detection:

1. **Test cross** is the most common method of detecting the linkage (**Figure 2**). In this method, the F1 heterozygous at two loci (say AaBb) is crossed to a double recessive parent (aabb) and the phenotypic ratio of test cross progeny is examined. If the phenotypie ratio of test cross progeny shows 1: 1 : 1 : 1 ratio of parental and recombinant genotypes, it indicates absence of linkage. If the frequency of parental types and recombinant types deviate significantly from the normal test cross ratio of 1 : 1 : 1 : 1, it

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

reveals presence of linkage between two genes under study. The proportion of the parental genotype would be relatively higher.

2. **Normal di-hybrid cross:** There is another way to detect the presence or absence of linkage. The individual heterozygous at two loci (AaBb) is self-fertilised. If there is complete dominance at each locus and no epistasis, the segregation ratio of the progeny will be 9: 3: 3 :1. Presence of linkage will lead to significant deviation from 9 : 3 : 3 : 1 ratio. The deviation of observed values from the expected ratio is tested with the help of χ^2 test. The proportion of the parental genotype would be relatively higher.

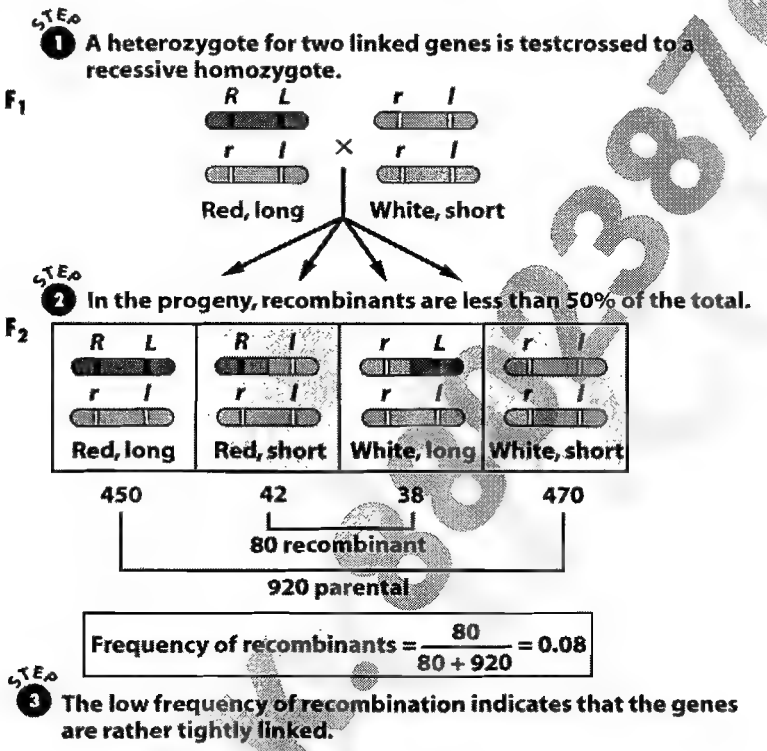


Figure 8: Test Cross Based Detecttion of Linkage

Consequences of Linkage

1. The biggest consequence of genetic linkage is the **preservation or predominance of the parental allelic combination** of genes in the subsequent generations of the organism. This may have both positive (transmission of some desirable trait) and negative (transmission of some deleterious trait, such as an allele for a disease) outcomes for the organism concerned. In all the cases, there is a statistical deviation from the independent assortment ratios (i.e. 9:3:3:1 for a di-hybrid cross and 1:1:1:1 for a test cross), as shown in the diagram on page 1 as well as in the figure above.

2. Another consequence of linkage is that **some traits show their differential expression based on sex** of the individual, because the gene controlling the trait happens to be sex-chromosome linked. Some well studied examples of sex linkage include eye colour in *Drosophila*, color blindness in humans, X-linked hemophilia in humans etc.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Linkage Mapping

Linkage mapping is a method of obtaining the relative distance and position of the genes on a chromosome based on the strength of linkage between different sets of genes. The strength is measured in terms of the incidences of crossing over between the genes.

Scientific Basis: The amount of crossing-over between various linked genes differs. This is an outcome of various genes located at different distances from one another on a chromosome. The basic rule is – the larger the distance between two genes on a chromosome, the greater would be the frequency of crossing over between them.

Alfred Sturtevant, a student working in the lab of **TH Morgan** in 1910s, developed a method for determining relative distance between linked genes based on the frequency of crossing over between them. Crossing over breaks linkage and gives rise to the rcombinants. Sturtevant postulated a rough proportionality: the greater the distance between the linked genes, the greater the chance that nonsister chromatids would cross over in the region between the genes and, hence, the greater the proportion of recombinants that would be produced. Thus, by determining the frequency of recombinants, we can obtain a measure of the distance between the genes. Sturtevant suggested that we can use the percentage of recombinants as a quantitative index of the linear distance between two genes on a genetic map, or **linkage map**.

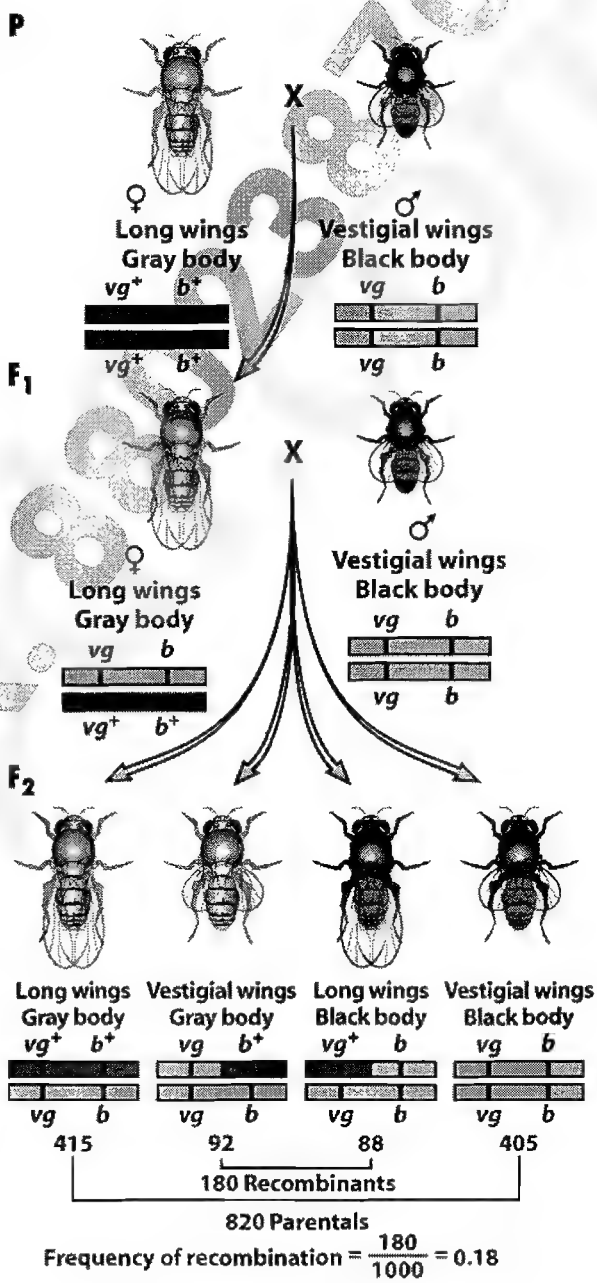


Figure 9: Operation of a Two Point Test Cross

We can define one **genetic map unit (m.u.)** as that distance between genes for which one product of meiosis out of 100 is recombinant. So, a **recombinant frequency (RF)** of 0.01 (1 percent) is defined as 1 m.u. [A map unit is sometimes referred to as a centimorgan (cM) in honor of Thomas Hunt Morgan.]

Practice: There are two types of linkage mapping.

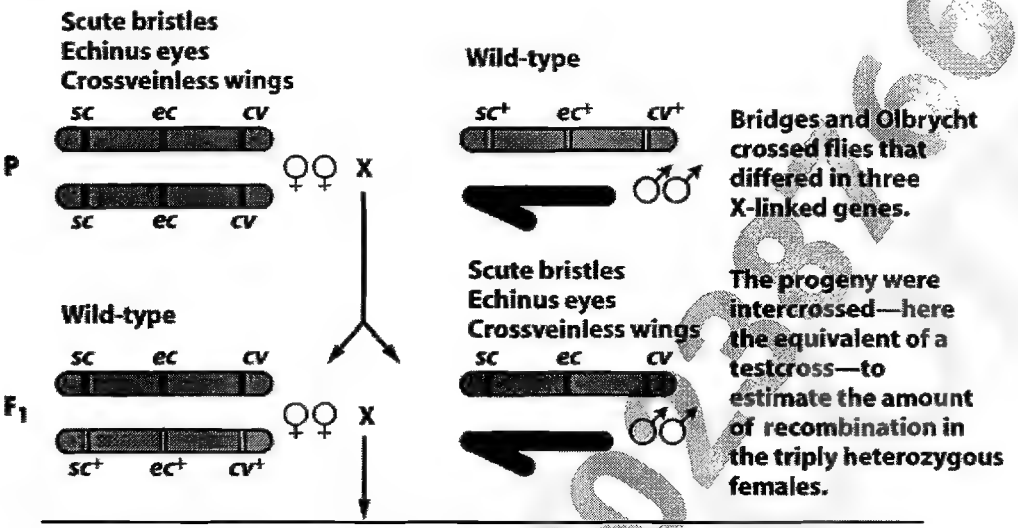
1. **Two point test cross:** It involves only two genes and it is used for determining the relative distance between two gene loci on the same chromosome. An experimental operation is shown in **Figure 3**. In the

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

illustrated example, the distance between the genes *vg* and *b* is 0.18 x 100 = 18 m.u. / cM (i.e. % frequency of recombination).

2. **Three point test cross:** This operation involves three genes and its purpose is to determine relative position of the genes on a chromosome. One experimental operation is shown below in **Figure 4**.



F_2

Class	Phenotype	Genotype of maternally inherited X chromosome			Number observed
1	Scute, echinus, crossveinless	*sc*	*ec*	*cv*	1158
2	Wild-type	*sc*+	*ec*+	*cv*+	1455
3	Scute	*sc*	*ec*+	*cv*+	163
4	Echinus, crossveinless	*sc*+	*ec*	*cv*	130
5	Scute, echinus	*sc*	*ec*	*cv*+	192
6	Crossveinless	*sc*+	*ec*+	*cv*	148
7	Scute, crossveinless	*sc*	*ec*+	*cv*	1
8	Echinus	*sc*+	*ec*	*cv*+	1
				Total:	3248

Figure 10: Operation of a Three Point Test Cross

This cross involves three genes, *sc*, *ec*, and *cv*.

Based on the cross results,

1. The number of cross-overs between the genes *sc* and *ec* is 295 (i.e. 163+130+1+1). Hence the % recombinant frequency is:

   (295 / 3248) x 100 = 9.1. Hence, the distance between the genes *sc* and *ec* is 9.1cM.

2. The number of cross-overs between the genes *sc* and *cv* is 637 (i.e. 163+130+192+148+2+2). Hence the % recombinant frequency is:

   (637 / 3248) x 100 = 19.6. Hence, the distance between the genes *sc* and *cv* is 19.6 cM.

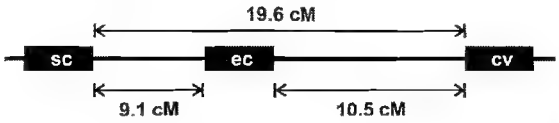
Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

3. The number of cross-overs between the genes *ec* and *cv* is 342 (i.e. 192+148+1+1). Hence the % recombinant frequency is:

   (342 / 3248) x 100 = 10.5. Hence, the distance between the genes *ec* and *cv* is 10.5cM.

Based on the above data, there can be only one order of the three genes on the X-chromosme of *Drosophila*, where they are located. This is:



Apart from the above arrangement, there can be no second arrangement that conforms to the available cross data. Thus, the three point test cross is a reliable method of determining the relative position of the genes on a chromosome.

Accuracy of linkage mapping: Despite a fair degree of accuracy, the method of linkage mapping is not 100% accurate. The scientific basis of this is that the crossing overs in the two regions are not completely independent of one another. In most of the experiments, there is some kind of ***interference***: that is, a crossover reduces the probability of a crossover in an adjacent region.

Interference is quantified by first calculating a term called the **coefficient of coincidence (c.o.c.),** which is the ratio of observed to expected double recombinants subtracted from 1. Hence

Interference (I) = 1 − c.o.c. =

$$1 - \left[\frac{\text{observed frequency, or number of double recombinants}}{\text{expected frequency, or number of double recombinants}} \right]$$

In some regions, there are never any observed double recombinants. In these cases, c.o.c. = 0, so I = 1 and interference is complete. Most of the time, the interference values that are encountered in mapping chromosome loci are between 0 and 1; but, in certain special situations, observed doubles exceed expected, giving negative interference values.

Moreover, at very long distances the linkage mapping does not give an accurate estimate.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 8: Crossing over

Introduction to Recombination

Recombination is a broad term in genetics that covers any molecular process, which causes separation of genes on the same chromosome and the occurrence of new gene combinations in progeny not seen in previous generations. In other words, recombination breaks the syntenic relation between the alleles of two different genes.

Apart from mutation, recombination is another important process that brings about genetic variation in populations. The main distinction between mutation and recombination is that while mutation creates new genetic information, recombination serves to re-arrange the existing genetic information.

The Basic Types of Recombination

There are following types of molecular processes, which bring about creation of new combination of genetic information.

1. **Homologous recombination**: A type of recombination that involves a genetic exchange between two similar or identical strands of DNA. Meiotic recombination (Crossing over) is an example of homologous recombination. This type of recombination often results in the exchange of alleles between chromosomes.

2. **Site-specific recombination**: This type of recombination involves the exchange of genetic material at very specific sites only. Examples include the integration of bacteriophage lambda into the host chromosome to form the prophage and the rearrangement of chromosomal DNA prior to expressing antibody genes.

3. **Transpositional Recombination**: It moves specialized nucleotide sequences, called *mobile genetic elements,* between nonhomologous sites within a genome. The movement can occur between two different positions in a single chromosome, as well as between two different chromosomes.

   Transpositional recombination can proceed via either of two distinct mechanisms, each of which requires specialized recombination enzymes and specific DNA sites.

   a. **Transposition by cut and paste** usually involves breakage reactions at the ends of the mobile DNA segments embedded in chromosomes and the attachment of those ends at one of many different nonhomologous target DNA sites. It does not involve the formation of heteroduplex DNA.

   b. **Conservative transpositional recombination** involves DNA Replication and the production of a very short heteroduplex joint.

4. **Illegitimate recombination**: There are a number of other genetic exchanges, which do not fall into any of the above classes - hence their name is illegitimate recombination. Examples include chromosomal fusion etc.

5. **Artificial recombination**: It is form of a synthetic DNA production that is made through the combination or insertion of one or more DNA strands, thereby combining DNA sequences that would not normally occur together. t differs from other methods of

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

genetic recombination, in the respect that that it does not occur through processes within the cell, but is exclusively engineered.

Crossing over & Its Cellular Basis

Fundamental Concepts

In eukaryotes, recombination results when crossing-over during meiosis separates linked genes. Crossing over is the physical process that the chromosomes participate in and result into the reshuffling of linked genes. Crossing over often occurs during prophase 1 of meiosis. Rarely, it also occurs during mitosis, as in *Drosophila*. In multicellular organisms, mitotic recombination can produce genetic mosaicism in which different cells have different genotypes.

Crossing over can be defined as the process by which two homologous chromosomes pair up and exchange sections of their DNA between non-sister chromatids. It results n the exchange of alleles between chromosomes but the order of the genes on the interacting chromosomes typically remains the same.

The History of Crossing Over Studies

- In 1909, the Belgian cytologist Frans Janssens described structures called chiasmata during prophase of the first meiotic division. He suggested that they represent regions in which nonsister chromatids of homologous chromosomes cross over each other
- In 1912, Thomas Hunt Morgan suggested that the chiasmata observed through the light microscope were sites of chromosome breakage and exchange resulting in genetic recombination.
- In 1931, Harriet Creighton and Barbara McClintock studied two loci of chromosome 9 of corn: one affecting seed color (*C*, colored; *c*, colorless) and the other affecting endosperm composition (*Wx*, waxy; *wx*, starchy). In this study, they correlated the genetic and cytological events of crossing over.

The Cellular Basis of Crossing Over

Reciprocal exchanges between non-sister chromatids of the homologous chromosomes are the physical or cellular basis of recombination.

Although Morgan's idea that the physical breaking and rejoining of chromosomes during meiosis was the basis of genetic recombination seemed acceptable, before 1930 no one had produced visible evidence that crossing-over between homologous chromosomes actually occurs.

In 1931, Harriet Creighton and Barbara McClintock, who studied corn, and Curt Stern, who worked with Drosophila, published the results of experiments showing that genetic recombination indeed depends on the reciprocal exchange of parts between maternal and paternal chromosomes. Stern, for example, bred female flies with two different X chromosomes, each containing a distinct physical marker near one of the ends.

Figure 1 outlines the currently known steps of recombination as they appear in chromosomes viewed through the light microscope during Meiosis I.

1. In Fig. 1a, the two homologs of each chromosome pair have already replicated, so there are now two pairs of sister chromatids or a total of four chromatids within each bivalent.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

2. In Fig. 1b, the synaptonemal complex zips together homologous chromosome pairs along their length. The synaptonemal zipper aligns homologous regions of all four chromatids such that allelic DNA sequences are physically near each other.

3. This proximity facilitates crossing-over between homologous sequences; as the biochemical mechanism of recombination requires a close interaction of DNAs on homologous chromosomes that have identical, or nearly identical, nucleotide sequences.

4. In Fig. 1c, the synaptonemal complex begins to disassemble. Although some steps of the recombination process occurred while the chromatids were zipped in synapsis, it is in this phase that the recombination event becomes apparent.

5. As the zipper dissolves, homologous chromosomes remain attached at chiasmata, the actual sites of crossing-over. Visible in the light microscope, chiasmata indicate where chromatid sections have switched from one molecule to another.

6. In Fig. 1d, during anaphase I, as the two homologs separate, starting at their centromeres, the ends of the two recombined chromatids pull free of their respective sister chromatids, and the chiasmata shift from their original positions toward a chromosome end, or telomere. This movement of chiasmata is known as terminalization. When the chiasmata reach the telomeres, the homologous chromosomes can separate from each other (Fig. 1e).

7. Meiosis continues and eventually produces four haploid cells that contain one chromatid—now a chromosome—apiece (Fig. 1f). Homologous chromosomes have exchanged parts.

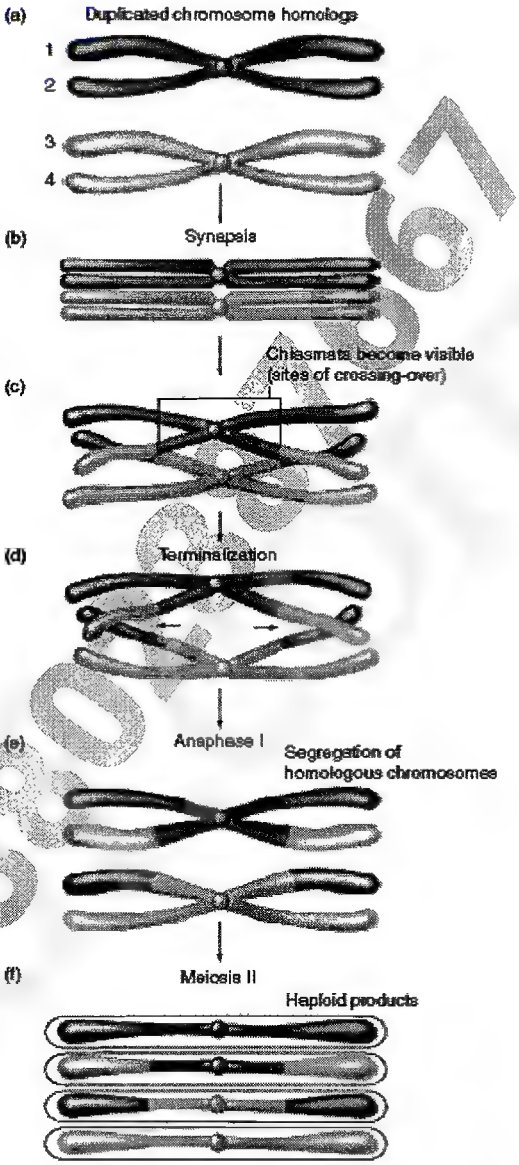


Figure 1: The cellular basis of recombination through the light microscope. (a) A pair of duplicated homologous chromosomes very early in prophase of meiosis I. (b) During leptotene and zygotene of prophase I, the synaptonemal complex helps align corresponding regions of homologous chromosomes, allowing recombination. (c) As the synaptonemal complex disassembles during diplotene, homologous chromosomes remain attached at chiasmata. (d) and (e) The chiasmata terminalize (move toward the chromosome ends), allowing the recombined chromosomes to separate during anaphase and telophase. (f) The result of the process is recombinant gametes.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Recombination can also take place during mitosis. It also occurs with the circular chromosomes of prokaryotic organisms and with cellular organelles such as mitochondria and chloroplasts, which do not undergo meiosis and do not form chiasmata.

The molecular mechanism of crossing over

The molecular mechanisms of the general recombination seem to be the same in all organisms (Alberts *et al,* 2008). Most information about the biochemistry of genetic recombination was originally derived from studies of *E. coli* and its viruses. Later, information was also obtained from experiments with simple eukaryotes such as yeasts. More recently, recombination mechanisms have extensively characterized in *Drosophila,* mice, and humans as well.

The general recombination observed in meiosis has the following characteristics:

1. Two homologous DNA molecules that were originally part of different chromosomes "cross over;" that is, their double helices break and the two broken ends join to their opposite partners to re-form two intact double helices

2. The site of exchange can occur anywhere in the homologous nucleotide sequences of the two participating DNA molecules.

3. General recombination often involves a ***Holliday Junction***. In a Holliday junction, the two homologous DNA helices that have initially paired are held together by the reciprocal exchange of two of the four strands present, one originating from each of the helices.

4. At the site of exchange, a strand of one DNA molecule has become base-paired to a strand of the second DNA molecule to create a *heteroduplex joint* that links the two double helices

5. No nucleotide sequences are altered at the site of exchange; some DNA replication usually takes place, but the cleavage and rejoining events occur so precisely that not a single nucleotide is lost or gained.

The current status of our understanding of Homologous Recombination is summarized in Figure 2.

Significance of Recombination

1. DNA sequences are rearranged by recombination mechanisms. The particular combination of genes present in any individual genome, as well as the timing and the level of expression of these genes, is often altered by such DNA rearrangements. In a population, this type of genetic variation is crucial to allow organisms to evolve in response to a changing environment. The evolutionary benefit of this type of gene mixing is apparently so great that the reassortment of genes by general recombination is not confined to multicellular organisms; it is also widespread in single-celled organisms.

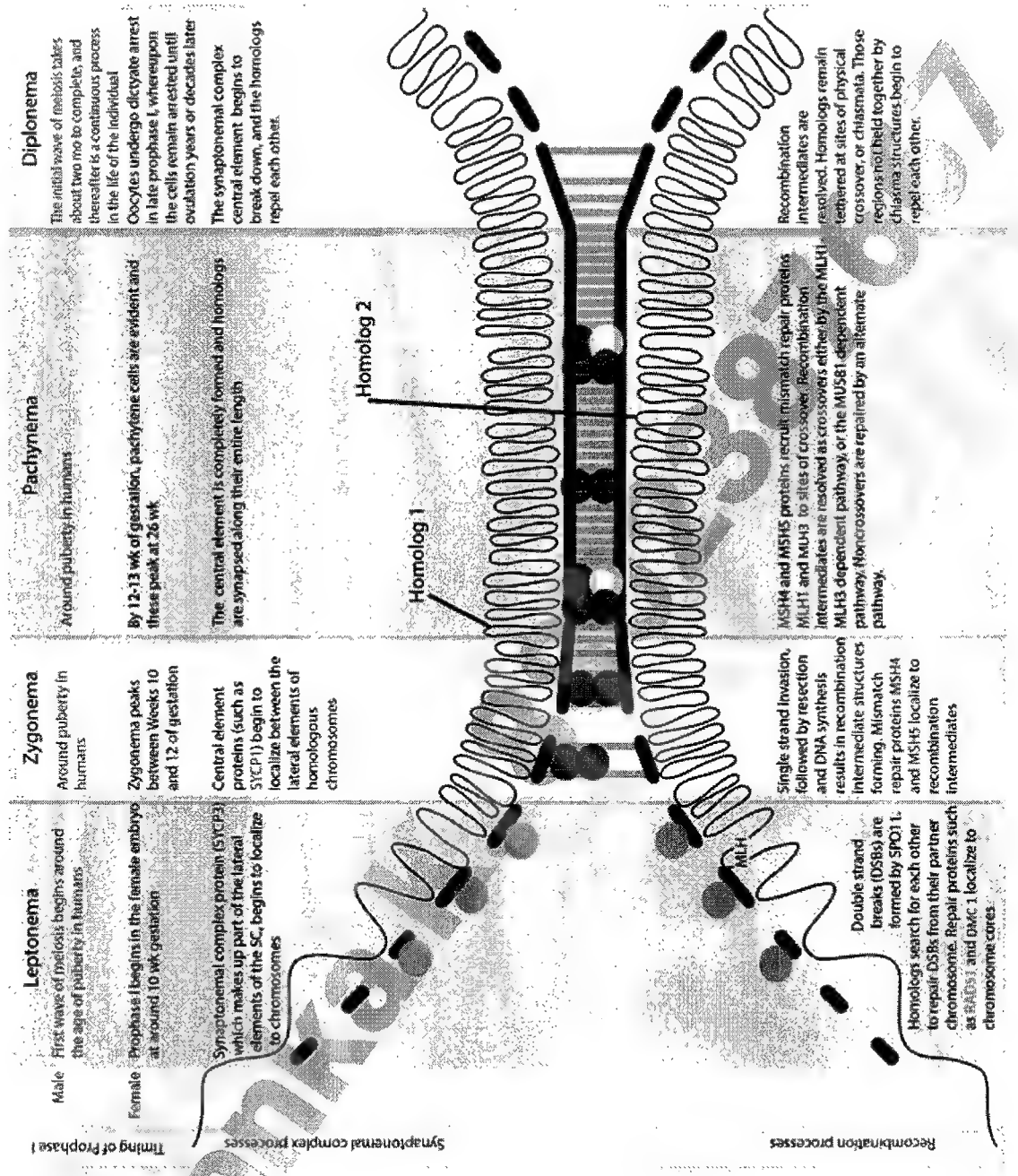
2. The general recombination reaction is essential for every proliferating cell, because accidents occur during nearly every round of DNA replication that interrupt the replication fork and require general recombination mechanisms to repair.

3. General recombination is also essential for the accurate chromosome segregation that occurs during meiosis in fungi, plants, and animals.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Figure 2: Molecular basis of crossing over



Priyanka 2017

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 9: Molecular Maps of chromosomes

Molecular mapping (also known as **Physical Mapping**) is a method of mapping genes or DNA sequences on a chromosome which does not rely on meiotic segregation (as done in *genetic mapping*).

Molecular mapping uses the tools and techniques of molecular biotechnology and provides us mapping data on genes and other gene sequences much more quickly and accurately than the genetic mapping methods. This explains why these methods are now increasingly being preferred by geneticists across the world. In all the genome projects, molecular mapping of chromosomes constitutes an important preliminary step before the sequencing of genomic DNA.

Basic types

A wide variety of molecular mapping strategies have been used to analyze the DNA of complex eukaryotic genomes. Mostly, an arbitrary distinction is made between two classes of molecular mapping:

- **low resolution molecular mapping** - the smallest map unit that can be resolved is typically one to several megabases (10^6) of DNA;
- **high resolution molecular mapping** - the resolution is typically very high, from hundreds of kilobases (10^3) to a single nucleotide.

Since complex eukaryotic DNA has only a very small percentage of coding DNA (~3% in the case of the human genome), a variety of **transcript mapping methods** have been developed recently for selectively identifying and studying transcribed sequences, which correspond to functional genes.

Important methods of molecular mapping

A variety of physical mapping techniques has been developed in the recent decades, the most important being:

1. **Restriction mapping**, which locates the relative positions on a DNA molecule of the recognition sequences for restriction endonucleases;
2. **Fluorescent*in situ*** **hybridization (FISH)**, in which marker locations are mapped by hybridizing a probe containing the marker to intact chromosomes;
3. **Sequence tagged site (STS) mapping,** in which the positions of short sequences are mapped by PCR and/or hybridization analysis of genome fragments.

Restriction mapping

Genetic mapping using ***Restriction fragment length polymorphisms*** (RFLPs) as DNA markers can locate the positions of polymorphic restriction sites within a genome. RFLPs were the first type of DNA marker to be studied.

The restriction enzymes cut DNA molecules at specific recognition sequences. This sequence specificity means that treatment of a DNA molecule with a restriction enzyme should always produce the same set of fragments. This is not always the case with genomic DNA molecules because some restriction sites are polymorphic due to mutations within the restriction sites, existing as two alleles, one allele displaying the correct sequence for the restriction site and

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

therefore being cut when the DNA is treated with the enzyme, and the second allele having a sequence alteration so the restriction site is no longer recognized. The result of the sequence alteration is that the two adjacent restriction fragments remain linked together after treatment with the enzyme, leading to a length polymorphism (*Figure 1*). This is an RFLP and its position on a genome map can be worked out by following the inheritance of its alleles, just as is done when genes are used as markers.

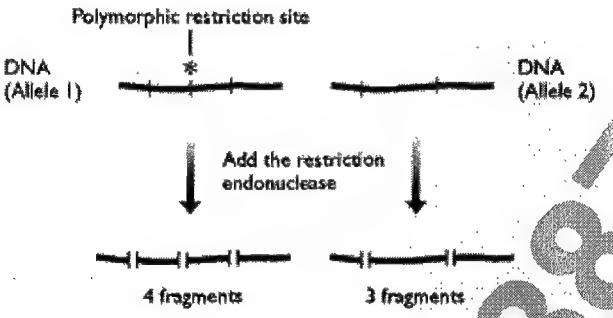


Figure 1: A restriction fragment length polymorphism (RFLP). The DNA molecule on the left has a polymorphic restriction site (marked with the asterisk) that is not present in the molecule on the right. The RFLP is revealed after treatment with the restriction enzyme because one of the molecules is cut into four fragments whereas the other is cut into three fragments.

There are thought to be about 10^5 RFLPs in the human or other complex genomes. In order to score an RFLP, it is necessary to determine the size of just one or two individual restriction fragments against a background of many irrelevant fragments.

It is also possible to use methods other than electrophoresis to map restriction sites in DNA molecules. With the technique called **optical mapping** (Schwartz*et al.*, 1993), restriction sites are directly located by looking at the cut DNA molecules with a microscope. The DNA must first be attached to a glass slide in such a way that the individual molecules become stretched out, rather than clumped together in a mass. There are two ways of doing this: gel stretching and molecular combing.

Fluorescent *in situ* hybridization (FISH)

In situ hybridization is a version of hybridization analysis in which an intact chromosome is examined by probing it with a labeled DNA molecule. The position on the chromosome at which hybridization occurs provides information about the map location of the DNA sequence used as the probe (Figure 2).

For the method to work, the DNA in the chromosome must be made single stranded ('denatured') by breaking the base pairs that hold the double helix together. Only then will the chromosomal DNA be able to hybridize with the probe. The standard method for denaturing chromosomal DNA without destroying the morphology of the chromosome is to dry the preparation onto a glass microscope slide and then treat with *formamide.*

In the early versions of *in situ* hybridization the probe was radioactively labeled but this procedure was unsatisfactory because it is difficult to achieve both sensitivity and resolution with a radioactive label, two critical requirements for successful *in situ* hybridization. Sensitivity requires that the radioactive label has a high emission energy (an example of such a radiolabel is ^{32}P), but if the radiolabel has a high emission energy then it scatters its signal and so gives poor resolution. High resolution is possible if a radiolabel with low emission energy, such as ^{3}H,

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

is used, but these have such low sensitivity that lengthy exposures are needed, leading to a high background and difficulties in discerning the genuine signal.

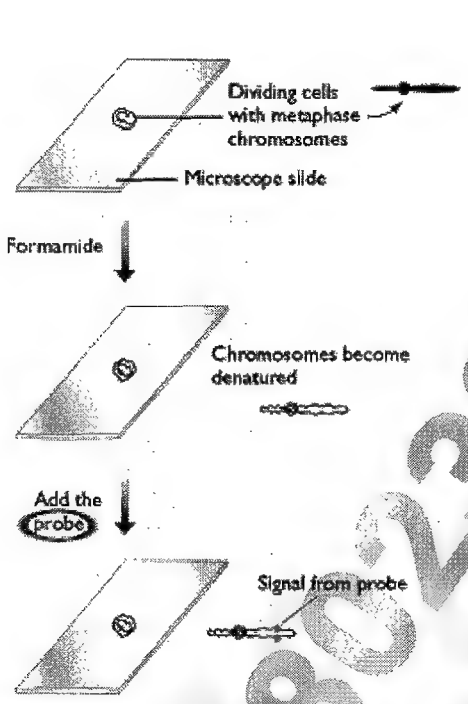


Figure 2: Fluorescent *in situ* hybridization. A sample of dividing cells is dried onto a microscope slide and treated with formamide so that the chromosomes become denatured but do not lose their characteristic metaphase morphologies. The position at which the probe hybridizes to the chromosomal DNA is visualized by detecting the fluorescent signal emitted by the labeled DNA.

These problems were solved in the late 1980s by the development of non-radioactive fluorescent DNA labels. These labels combine high sensitivity with high resolution and are ideal for *in situ* hybridization. Fluorolabels with different colored emissions have been designed, making it possible to hybridize a number of different probes to a single chromosome and distinguish their individual hybridization signals, thus enabling the relative positions of the probe sequences to be mapped. To maximize sensitivity, the probes must be labeled as heavily as possible.

As far as the construction of a physical map is concerned, a cloned DNA fragment can be looked upon as simply another type of marker. Mapping the positions of clones therefore provides a direct link between a genome map and its DNA sequence.

If the probe is a long fragment of DNA then one potential problem, at least with higher eukaryotes, is that it is likely to contain examples of repetitive DNA sequences and so may hybridize to many chromosomal positions, not just the specific point to which it is perfectly matched. To reduce this non-specific hybridization, the probe, before use, is mixed with unlabeled DNA from the organism being studied. The idea is that the unlabeled DNA hybridizes to the repetitive DNA sequences in the probe, blocking these so that the subsequent *in situ* hybridization is driven wholly by the unique sequences (Lichter *et al.*, 1990). Non-specific hybridization is therefore reduced or eliminated entirely.

Sequence tagged site (STS) mapping

To generate a detailed physical map of a large genome we need, ideally, a high-resolution mapping procedure that is rapid and not technically demanding. Neither of restriction mapping

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

and FISH meets these requirements. Restriction mapping is rapid, easy, and provides detailed information, but it cannot be applied to large genomes. FISH can be applied to large genomes, and modified versions such as fiber-FISH can give high-resolution data, but FISH is difficult to carry out and data accumulation is slow, map positions for no more than three or four markers being obtained in a single experiment.

At present the most powerful physical mapping technique, and the one that has been responsible for generation of the most detailed maps of large genomes, is STS mapping. A sequence tagged site or **STS** is simply a short DNA sequence, generally between 100 and 500 bp in length, that is easily recognizable and occurs only once in the chromosome or genome being studied. To map a set of STSs, a collection of overlapping DNA fragments from a single chromosome or from the entire genome is needed. In the example shown in *Figure 3,* a fragment collection has been prepared from a single chromosome, with each point along the chromosome represented on average five times in the collection.

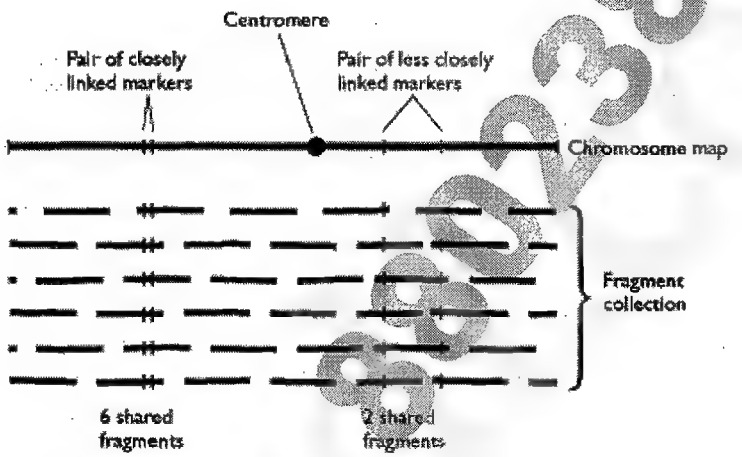


Figure 3: STS mapping. The fragments span the entire length of a chromosome, with each point on the chromosome present in an average of five fragments. The two blue markers are close together on the chromosome map and there is a high probability that they will be found on the same fragment. The two green markers are more distant from one another and so are less likely to be found on the same fragment.

The data from which the map will be derived are obtained by determining which fragments contain which STSs. This can be done by hybridization analysis but PCR is generally used because it is quicker and has proven to be more amenable to automation. The chances of two STSs being present on the same fragment will, of course, depend on how close together they are in the genome. If they are very close then there is a good chance that they will always be on the same fragment; if they are further apart then sometimes they will be on the same fragment and sometimes they will not. The data can therefore be used to calculate the distance between two markers, in a manner analogous to the way in which map distances are determined by linkage analysis. In linkage analysis a map distance is calculated from the frequency at which crossovers occur between two markers. STS mapping is essentially the same, except that each map distance is based on the frequency at which *breaks* occur between two markers.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 10: Sex chromosomes and sex determination

What Is Sex Determination?

Sex of an individual refers to what type/s of gamete it would make during sexual reproduction. Organisms can show the following types of sexes:

1. **Male** – produces the sperms
2. **Female** – produces the eggs
3. **Hermaphrodite** – produces both sperms and eggs.
4. **Intersex** – produces no viable gamete, but morphologically it is an intermediate between the male and the female

Sex determination is the organism's commitment to go through the developmental process that would make it attain a particular type of sexual phenotype. This determination occurs mostly by genetic methods, yet in some organisms physiological, environmental or social factors may play a role.

What Is Sex Differentiation?

This is the developmental process following the organism's sex determination commitment that ultimately gives rise to the sexual phenotype of the organism. There are **two levels of sex differentiation in complex organisms**:

1. **Primary sex differentiation** that results into the formation of the gonads, which ultimately give rise to the gametes.
2. **Secondary sex differentiation** forms the structures not directly responsible for gamete production, but which otherwise help in the sexual process.

Depending on the organism, the primary sexual differentiation can be one of the following **two basic types**:

1. **An individual can have only one sex – male or female**: This situation is called **Gonochoric** in animals and **Dioecious** in plants.
2. **An individual can have both the sexes**: This situation is called **Hermaphroditic** in animals and **Monoecious** in plants.

A third type referred to as **intersex** is seen is some organisms, as in *Drosophila*. In such cases, phenotype overlaps between male and female but there is no production of functional gametes.

How is Sex Determined?

There are many ways in which sex differences arise.

- In many cases, sex determination is **genic**: males and females

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

have different alleles or even different genes that specify their sexual morphology.

- In animals, sex determination is mostly **chromosome controlled**, sex chromosomes or otherwise.
- In other cases, sex is **determined by environmental variables** (such as temperature, locations etc.)
- **Social variables** (the size of an organism relative to other members of its population) may also control the sex of the organism.
- The details of some sex-determination systems are not yet fully understood.

Genic Control of Sex

In some plants, fungi and protistans, sex is genetically determined, but there are no obvious differences in the chromosomes of males and females—hence, there are no sex chromosomes. These organisms have **genic sex determination**; genotypes at one or more loci determine the sex of an individual plant, fungus and protistan.

Well studied example includes the green alga *Chlamydomonas*, the fungi *Saccharomyces* and *Neurospora* and the protozoa *Tetrahymena*. Among the plants, it is seen in Papaya and *Ecballium elaterium*.

Here, it is important to realize that, even in chromosomal sex-determining systems, sex is actually determined by in genes.

Chromosomal Control of Sex

The chromosomes control sex of the individual in several ways.

1. By Sex Chromosome based system
2. By Ploidy Variations
3. By Autosome : Sex Chromosome ratio

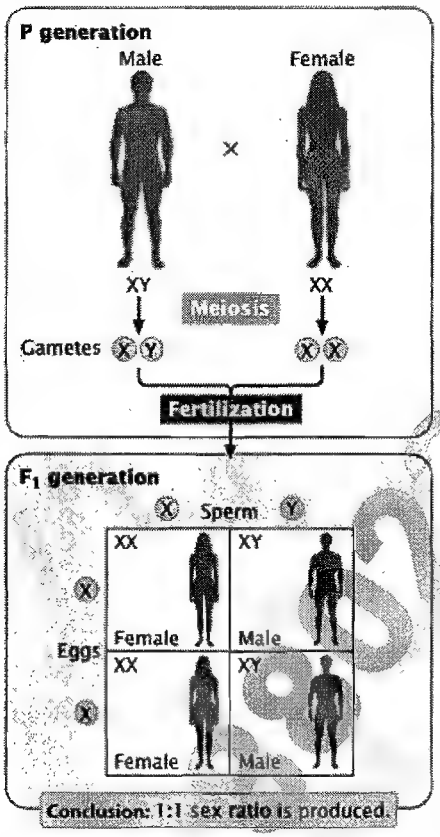
THE SEX CHROMOSME BASED SYSTEM

1. The XX-XY sex-determination system is a well-known sex-determination system, found in human beings, other mammals and some insects like *Lygaeus turicus*. The XX-XY sex determination system was first described independently by **Dr. Nettie Stevens** and **Edmund Beecher Wilson** in 1907 in the Hemipteran insect *Lygaeus turicus*. This sex determination system is also called **Lygaeus Type**. In the XY sex-determination system, females have

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

two of the same kind of sex chromosome (XX), while males have two distinct sex chromosomes (XY).



During gametogenesis, the male produces two types of gametes, X and Y – while the female partner produces only one type of gamete, i.e. X. Post fertilization, the XY zygote develops as Male while the XX zygote develops as Female. Because males produce two different types of gametes with respect to the sex chromosomes, they are said to be the **heterogametic sex**. Females, which produce gametes that are all the same with respect to the sex chromosomes, are the **homogametic sex**.

Among the plants, this system is found in *Melandrium album* (Now known as *Sinapis alba*), *Coccinia indica* and *Spinacea oleracea*.

2. The XX-XO sex-determination system was the first sex-determination system to be known by EB Wilson in 1906, in the Orthopteran insect *Protenor*. It is also known as **Protenor Type** of sex-determination system. Subsequently, this type of sex-determination was also reported from Grasshoppers, Roaches, *Anasa* and the plant *Vallisnaria*.

In this sex-determination system, Edmund B. Wilson demonstrated that female somatic cells in the insect *Protenor* contain 14 chromosomes, including 2 X chromosomes. During oogenesis, an even reduction occurs, producing gametes with 7 chromosomes, including one X. Male somatic cells, on the other hand, contain only 13 chromosomes, including a single X chromosome. During spermatogenesis, gametes are produced containing either 6 chromosomes, without an X, or 7 chromosomes, one of which is an X. Fertilization by X-bearing sperm results in female offspring, and fertilization by X-deficient sperm results in male offspring.

Because males produce two different types of gametes with respect to the sex chromosomes, they are said to be the **heterogametic sex**. Females, which produce gametes that are all the same with respect

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

to the sex chromosomes, are the **homogametic sex**.

3. ZZ-ZW sex determination In this system, the female is heterogametic and the male is homogametic. To prevent confusion with the XX-XY system, the sex chromosomes in this system are labeled Z and W. Females in this system are ZW; after meiosis, half of the eggs have a Z chromosome and the other half have a W. Males are ZZ; all sperm contain a single Z chromosome. The ZZ-ZW system is found in birds, also in all snake species, and some other reptiles, some amphibians, and some fishes like *Carassius carassius*, *Negaprion* and *Etmopterus*.

4. ZZ-ZO sex determination In this system, which is operationally opposite to the XX-XO sex-determination system, the female is homogametic and the male is heterogametic. To prevent confusion with the XX-XO system, the sex chromosomes in this system are labeled Z. Males in this system are ZZ; after meiosis, all the sperms have a Z chromosome. Females are ZO; all eggs contain either a single Z chromosome or no sex chromosome at all. The ZZ-ZO system is found in certain moths, and some butterflies.

5. Compound Sex Chromosome based sex determination

Here, a compound of multiple chromosomes generates the effect of single sex chromosome. If any one of the chromosomes from the compound is missing, the effect will be of complete absence of the sex chromosome.

It has three subtypes.

a. **Compound X Phenomenon.** It is best studied in Nematode *Ascarias incurve*, where 8 chromosomes taken together give the effect of a single X-chromosome. Hence the genotype composition are:

Female → 2A : 8X + 8X; Male → 2A : 8X + Y

b. **Compound Y Phenomenon.** It is not reported in animals so far, but quiet frequent in higher plants. *Rumex haustaulus* is the best-studied example, where there is a compound of 2Y chromosome.

c. **Compound of Both X and Y.** It is found in both animals and plants; beetles *Blaps polychresta* is the well studied example. The plant *Humulus lupulus* also shows this phenomenon. In these cases, both X and Y chromosomes are present in compound of multiple chromosome.

SEX DETERMINATION BY PLOIDY VARIATIONS (HAPLODIPLOID SEX-DETERMINATION SYSTEM)

The **haplodiploid sex-determination system** determines the sex of the offspring of many Hymenopterans (bees, ants, and wasps), and

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

coleopterans (bark beetles). In this system, sex is determined by the number of sets of chromosomes an individual receives. An offspring formed from the union of a sperm and an egg develops as a female, and an unfertilized egg develops as a male. This means that the males have half the number of chromosomes that a female has, and are haploid. This system produces a number of peculiarities; chief among these is that a male has no father and cannot have sons, but he has a grandfather and can have grandsons.

The statement that Hymenoptera females are diploid and males haploid is a cytological description that requires a detailed genetic explanation. Five genetic hypotheses of sex determination have been presented, of which the explanation given by Whiting is most satisfactory.

The Theory of Involvement of a Single Series of Multiple Alleles (Whiting, 1943, 1945; Mackensen, 1951) Whiting based this on the studies of diploid males of *Bracon hebetor* (published as *Habrobracon juglandis*). From his crosses, he concluded that sex in this Braconid is determined by a series of at least 8 heteroalleles that in hemizygous (xo^1, xo^2... xo^8) or homozygous (xo^1xo^1, xo^2xo^2... xo^8xo^8) condition are males and in heterozygosis (xo^1xo^2,... xo^7xo^8) are females.

Mackensen (1951), using inbred lines of Honey bee (*Apis mellifera*), suggested that the same multiple alleles occurred in *Apis mellifera*. Woyke (1963, 1979, 1986) actually demonstrated that in *Apis mellifera* and *Apis cerana* the homozygous diploids for the sex alleles were males.

Chromosomal Sex Determination in *Drosophila*

In *Drosophila*, the chromosomal basis of sex determination is due to the ratio of X chromosomes to sets of autosomes. This mechanism, worked out by Calvin Bridges (1921, 1925) is known as **Genic Balance**.

In *Drosophila*, $n = 4$: one sex chromosome and three different autosomes. Hence, one autosomal set of three chromosomes is represented as A and, in a diploid fly, A = 2.

The effect of X:A ratio determines sex in *Drosophila* because sex determination in this species requires a balance of female determinants on the X chromosome and male determinants on the autosomes.

A normal 2X *Drosophila* diploid (XX AA) has an X:A ratio of 1.0 and is phenotypically female. An XY diploid (XY AA) has an X:A ratio of 0.5 and is male; an XO diploid also is male (although sterile). Triploids with three X chromosomes (XXX AAA) are females, those with one X (XYY AAA) are male, and those with two X's (XXY AAA)

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

are "in between" (intersexes).

As, XO *Drosophila* are sterile males, it becomes established that in flies, the Y chromosome is not involved in determining sex. Rather, it contains genes active in forming sperm in adults.

The table below shows the different X-to-autosome ratios and the resulting sex.

Table: Ratios of X chromosomes to autosomes in different sexual phenotypes in *Drosophila melanogaster*

X chromosomes	Autosome sets (A)	X:A ratio	Sex
3	2	1.50	Metafemale
4	3	1.33	Metafemale
4	4	1.00	Normal female
3	3	1.00	Normal female
2	2	1.00	Normal female
2	3	0.66	Intersex
1	2	0.50	Normal male
1	3	0.33	Metamale

The table above establishes the basic scheme of sex determination in *Drosophila* by X:A ratio. This scheme is:

1. X/A = 1.0—female
2. X/A = 0.5—male
3. X/A > 1.0—metafemale
4. X/A < 0.5—metamale
5. X/A between 0.5 and 1—intersex

Several genes with roles in sex determination have been found with the following observations.

1. Loss-of-function mutations in most of these genes—*Sex-*

Direct PDF https://t.me/CivilServices_UPSC

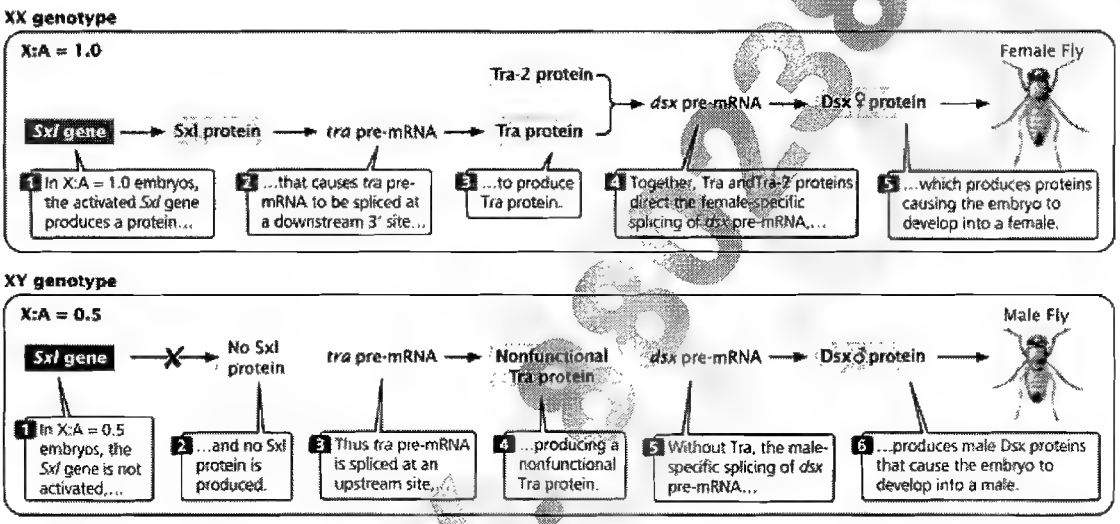
https://t.me/Upsc_Pdf

lethal (*Sxl*), *transformer* (*tra*), and *transformer-2* (*tra2*)—transform XX individuals into males. Such mutations have no effect on sex determination in XY males.

2. Homozygosity of the *intersex* (*ix*) gene causes XX flies to develop an intersex phenotype having portions of male and female tissue in the same organ.

3. he *doublesex* (*dsx*) gene is important for the sexual differentiation of both sexes. If *dsx* is absent, both XX and XY flies turn into intersexes.

Recent studies have generated the model of the regulatory cascade seen in *Drosophila* leading to sex determination.



Molecular basis of sex determination in *Drosophila*

Alternative mRNA splicing regulates the expression of genes which control sex of a fruit fly. Sex differentiation in *Drosophila* arises from a cascade of gene regulation. When the ratio of X chromosomes to the number of haploid sets of autosomes (the X: A ratio) is 1, a female-specific promoter is activated early in development and stimulates the transcription of the *sex-lethal (Sxl)* gene. The protein encoded by *Sxl* regulates the splicing of the pre-mRNA transcribed from another gene called *transformer (tra).*

The splicing of *tra* pre-mRNA results in the production of Tra protein. Together with another protein (Tra-2), Tra stimulates the female-specific splicing of pre-mRNA from yet another gene called *doublesex (dsx).* This event produces a female-specific Dsx protein, which causes the embryo to develop female characteristics.

In male embryos, which have an X:A ratio of 0.5, the promoter that transcribes the *Sxl* gene in females is inactive; so no Sxl protein is produced. In the absence of Sxl protein, Tra pre-mRNA is spliced at a different 3' splice site to produce a nonfunctional form of Tra

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

protein. In turn, the presence of this nonfunctional Tra in males causes Dsx pre-mRNAs to be spliced differently, and a male-specific Dsx protein is produced. This event causes the development of male-specific traits.

In summary, the Tra, Tra-2, and Sxl proteins regulate alternative splicing that produces male and female phenotypes in *Drosophila*. Exactly how these proteins regulate alternative splicing is not yet known.

Chromosomal Sex Determination in humans

In mammals, primary sex determination (that is the determination of the gonad) is strictly chromosomal.

The female is XX and the male is XY. Thus, every individual must have at least one X chromosome.

Since the female is XX, each of her eggs has a single X chromosome. The male, being XY, can generate two types of sperm: half bear the X chromosome, half the Y. If the egg receives another X chromosome from the sperm, the resulting individual is XX, forms ovaries, and is female; if the egg receives a Y chromosome from the sperm, the individual is XY, forms testes, and is male.

The Y chromosome, which is the fundamental determinant of sex in humans, carries a gene that encodes a testis-determining factor. This factor organizes the gonad into a testis rather than an ovary. The Y chromosome is a crucial factor for determining sex in mammals as its presence makes a male and its absence makes a female.

A person with five X chromosomes and one Y chromosome (XXXXXY) would be male. Furthermore, an individual with only a single X chromosome and no second X or Y (i.e., XO) develops as a female and begins making ovaries, although the ovarian follicles cannot be maintained. For a complete ovary, a second X chromosome is needed.

Secondary sex determination affects the bodily phenotype outside the gonads and these secondary sex characteristics are usually determined by hormones secreted from the gonads. However, in the absence of gonads, the female phenotype is generated. When Jost (1953) removed fetal rabbit gonads before they had differentiated, the resulting rabbits had a female phenotype, regardless of whether they were XX or XY. They each had oviducts, a uterus, and a vagina.

As mentioned earlier, the general scheme of mammalian sex determination is based on presence or absence of Y chromosome.

1. If the Y chromosome is absent, the gonadal primordia develop into ovaries. The ovaries produce **estrogen**, a hormone that enables the development of the **Müllerian**

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

duct into the uterus, oviducts, and upper end of the vagina.

2. If the Y chromosome is present, testes form and secrete two major hormones.

   a. The first hormone—**anti-Müllerian duct hormone** (**AMH**; also referred to as Müllerian-inhibiting substance, MIS)—destroys the Müllerian duct.

   b. The second hormone—**testosterone**—masculinizes the fetus, stimulating the formation of the penis, scrotum, and other portions of the male anatomy, as well as inhibiting the development of the breast primordia. Thus, the body has the female phenotype unless it is changed by the two hormones secreted by the fetal testes.

Several genes have been found whose function is necessary for normal sexual differentiation.

1. **Sry: the Y chromosome sex determinant:** The *Sry* gene (***sex-determining region of the Y chromosome***) has also been called the testis-determining factor on the Y chromosome (*TDF* in humans, *Tdy* in mice). In humans, this major gene for the testis-determining factor resides on the short arm of the Y chromosome, located near the tip occupying a 35,000-base-pair region. It encodes a peptide of 223 amino acids. This peptide is a transcription factor, since it contains a DNA-binding domain called the **HMG** (*h*igh-*m*obility *g*roup) **box**. It is thought that several testis-specific genes contain SRY-binding sites in their promoters or enhancers, and that the binding of SRY to these sites begins the developmental pathway to testis formation (Cohen et al. 1994).

2. **Sox9:** One of the autosomal genes involved in sex determination is *SOX9*, which encodes a putative transcription factor that also contains an HMG box. XX humans who have an extra copy of *SOX9* develop as males, even though they have no *SRY* gene. It is thus an autosomal sex reversal gene.

3. **Sf1:** It is the link between *Sry* and the male developmental pathways. It may be directly or indirectly activated by SRY. *Sf1* is necessary to make the bipotential gonad; but while Sf1 levels decline in the genital ridge of XX mouse embryos, the *Sf1* gene stays on in the developing testis. Sf1 appears to be active in masculinizing both the Leydig and the Sertoli cells. In the Sertoli cells, Sf1, working in collaboration with Sox9, is needed to elevate the levels of AMH transcription.

4. **Dax1:** In 1980, Bernstein and her colleagues reported two sisters who were genetically XY. Their Y chromosomes were normal, but they had a duplication of a small portion of the

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

short arm of the X chromosome. Subsequent cases were found, and it was concluded that if there were two copies of this region on the active X chromosome, the SRY signal would be reversed. This region contains a gene for a protein that competes with the SRY factor and it is important in directing the development of the ovary. In testicular development, this gene would be suppressed, but having two active copies of the gene would override this suppression. This gene, *DAX1*, has been cloned and shown to encode a member of the nuclear hormone receptor family. It is thus a potential ovary-determining gene on the X chromosome.

5. **Wnt4:** The *WNT4* gene is another gene that may be critical in ovary determination. This gene is expressed in the mouse genital ridge while it is still in its bipotential stage. *Wnt4* expression then becomes undetectable in XY gonads (which become testes), whereas it is maintained in XX gonads as they begin to form ovaries. In transgenic XX mice that lack the *Wnt4* genes, the ovary fails to form properly, and its cells express testis-specific markers, including AMH- and testosterone-producing enzymes.

Dosage compensation

The presence of different numbers of X chromosomes in males and females presents a special problem in development. Because females have two copies of every X-linked gene and males possess one copy, the amount of gene product (protein) from X-linked genes would normally differ in the two sexes — females would produce twice as much gene product as males. This difference could be highly detrimental because protein concentration plays a critical role in development.

Animals overcome this potential problem through dosage compensation, which equalizes the amount of protein produced by X-linked genes in the two sexes.

In fruit flies, dosage compensation is achieved by a doubling of the activity of the genes on the X chromosome of the male. In the worm *Caenorhabditis elegans,* it is achieved by a halving of the activity of genes on both of the X chromosomes in the female.

Placental mammals use yet another mechanism of dosage compensation; genes on one of the X chromosomes in the female are completely inactivated.

In 1949, Murray Barr observed condensed, darkly staining bodies in the nuclei of cells from female cats; this darkly staining structure became known as a Barr body.

Mary Lyon proposed in 1961 that the Barr body was an inactive X chromosome; her hypothesis (now proved) has become known as the

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Lyon hypothesis. She suggested that, within each female cell, one of the two X chromosomes becomes inactive; which X chromosome is inactivated is random.

The 3 main points of Lyon Hypothesis are as follows.

1. One X chromosome in each female cell becomes inactivated early in development and form a Barr Body;
2. Which one is deactivated is random;
3. All cells derived from cell with deactivated chromosome will have that chromosome deactivated.

If a cell contains more than two X chromosomes, all but one of them is inactivated.

Number of Barr Bodies:

1. XYAA—0 Barr bodies
2. XXAA—1 Barr body
3. X0AA—0 Barr bodies
4. XXXAA—2 Barr bodies
5. XXYAA—1 Barr body

As a result of X inactivation, females are functionally hemizygous at the cellular level for X-linked genes. In females that are heterozygous at an X-linked locus, approximately 50% of the cells will express one allele and 50% will express the other allele; thus, in heterozygous females, proteins encoded by both alleles are produced, although not within the same cell. This functional hemizygosity means that cells in females are not identical with respect to the expression of the genes on the X chromosome; females are mosaics for the expression of X-linked genes.

X inactivation takes place relatively early in development—in humans, within the first few weeks of development. Once an X chromosome becomes inactive in a cell, it remains inactivated and is inactive in all somatic cells that descend from the cell. Thus, neighboring cells tend to have the same X chromosome inactivated, producing a patchy pattern (mosaic) for the expression of an X-linked characteristic in heterozygous females.

Lyon's hypothesis suggests that the presence of variable numbers of X chromosomes should not be detrimental in mammals, because any X chromosomes beyond one should be inactivated. However, persons with Turner syndrome (XO) differ from normal females, and those with Klinefelter syndrome (XXY) differ from normal males. How do these conditions arise in the face of dosage compensation? The reason may lie partly in the fact that there is a short period of time,

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

very early in development, when all X chromosomes are active. If the number of X chromosomes is abnormal, any X-linked genes expressed during this early period will produce abnormal levels of gene product. Furthermore, the phenotypic abnormalities may arise because some X-linked genes escape inactivation.

Exactly how an X chromosome becomes inactivated is not completely understood, but it involves the addition of methyl groups ($-CH_3$) to the DNA. The *XIST* (for X inactive-specific transcript) gene, located on the X chromosome, is required for inactivation. Only the copy of *XIST* on the inactivated X chromosome is expressed, and it continues to be expressed during inactivation (unlike most other genes on the inactivated X chromosome). Interestingly, *XIST* does not encode a protein; it produces an RNA molecule that binds to the inactivated X chromosome. This binding is thought to prevent the attachment of other proteins that participate in transcription and, in this way, it brings about X inactivation

Anomalies related to Sex Determination in Humans

Hermaphrodites

1. Have both ovaries and testes
2. External genitalia are ambiguous
3. Generally, true hermaphrodites are sterile, but not always
4. In 1978 (at The American Society of Human Genetics meeting in Vancouver, B.C.) as case was reported in which a 25-year-old hermaphrodite delivered a stillborn child after 30 weeks gestation.
5. True hermaphrodites are genetic mosaics—some cells are XX, while others are XY.

Pseudohermaphrodites

1. Pseudohermaphrodites have either testicular or ovarian tissue, but not both
2. Generally the tissue is rudimentary
3. External genitalia are often ambiguous
4. Some are genetically female, but may look like males
5. Some are genetically male, but may look like females and lead normal female sex lives

Turner's Syndrome in humans (45, XO)

1. Due to nondisjunction in male or female parents to produce gametes without X chromosome

Priyanka 8302387

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

2. Incidence: 1/5000 female births
3. High proportion of spontaneous abortions are Turner's and most Turner's individuals are aborted
4. Phenotypic features:
    a. Short stature
    b. Webbed neck—web of skin between neck and shoulders
    c. Breast development absent or nearly so
    d. Some cognitive functions affected, but intelligence often about normal
    e. Pubic and axillary hair reduced or absent
    f. Infantile genitalia
    g. Usually sterile

Kleinfelter's Syndrome (47, XXY)

1. Due to nondisjunction of X chromosome in male or female parent
2. Incidence: 1/1000 male births
3. Tends to be maternal age effect
4. Sometimes have more than two X chromosomes
5. Phenotypic features
    a. Long arms
    b. Breast development
    c. Little or no sperm production
    d. Small testes
    e. Usually mentally retarded

XYY Condition

1. Due to nondisjunction of Y chromosome in male
2. Incidence: 1/1000 male births
3. Phenotypic features:
    a. Above average height
    b. Fertile

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

c. Sometimes (but not always) retarded

d. May be correlation with delinquency

Poly-X Females (XXX, XXXX, XXXXX,...)

1. Incidence: 1/1000 female births
2. Maternal age effect
3. Phenotypic features:
   a. Sometimes infantile genitalia
   b. Sometimes underdeveloped breasts
   c. Fertile
   d. Sometimes mental retardation
   e. Incidence increases with increasing number of X chromosomes

Environmental Sex Determination

Temperature-dependent sex determination in reptiles

While the sex of most snakes and most lizards is determined by sex chromosomes at the time of fertilization, the sex of most turtles and all species of crocodilians is determined by the environment after fertilization. In these reptiles, the temperature of the eggs during a certain period of development is the deciding factor in determining sex, and small changes in temperature can cause dramatic changes in the sex ratio (Bull 1980). Often, eggs incubated at low temperatures (22–27°C) produce one sex, whereas eggs incubated at higher temperatures (30°C and above) produce the other. There is only a small range of temperatures that permits both males and females to hatch from the same brood of eggs.

1. The eggs of the snapping turtle ***Macroclemys*** become female at either cool (22°C or lower) or hot (28°C or above) temperatures. Between these extremes, males predominate.
2. In laboratory studies, incubating ***Emys obicularis*** eggs at temperatures above 30°C produces all females, while temperatures below 25°C produce all-male broods. The threshold temperature (at which the sex ratio is even) is 28.5°C.
3. In crocodiles, in which temperature extremes produce females while moderate temperatures produce males.
4. Lizard *Agama agama* produces only males at temperature exceeding 26^0 C.

It appears that the enzyme **aromatase** (which can convert testosterone into estrogen) is important in temperature dependent

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

sex determination.

Chemical-dependent sex determination in frogs

Polychlorinated Biphenyl Compound (PCB), which pollutes water, leads to extensive feminization of Amphibian larvae. At a late stage, if the male frog is exposed to PCB it shows reproductive abnormality. Similarly, herbicide Atrazine causes feminization of frog tadpole.

Location-dependent sex determination

Location-dependent sex determination is well studied in *Bonellia* and *Crepidula.*

The sex of the worm *Bonellia* depends on where a larva settles. If a *Bonellia* larva lands on the ocean floor, it develops into a 10-cm-long female. If the larva is attracted to a female's proboscis, it travels along the tube until it enters the female's body. Therein it differentiates into a minute (1–3-mm-long) male that is essentially a sperm-producing symbiont of the female.

One fascinating example of environmental sex determination is seen in the marine mollusk *Crepidula fornicata*, also known as the common slipper limpet. Slipper limpets live in stacks, one on top of another. Each limpet begins life as a swimming larva. The first larva to settle on a solid, unoccupied substrate develops into a female limpet. It then produces chemicals that attract other larvae, which settle on top of it. These larvae develop into males, which then serve as mates for the limpet below. After a period of time, the males on top develop into females and, in turn, attract additional larvae that settle on top of the stack, develop into males, and serve as mates for the limpets under them. Limpets can form stacks of a dozen or more animals; the uppermost animals are always male. This type of sexual development is called **sequential hermaphroditism**; each individual animal can be both male and female, although not at the same time. In Crepidula fornicata, sex is determined environmentally by the limpet's position in the stack.

Social context-dependent sex determination

The sex of the blue-headed wrasse (*Thalassoma bifasciatum*) depends on the other fish it encounters. If the wrasse larva reaches a reef where a male lives with many females, it develops into a female. When the male dies, one of the females (usually the largest) becomes a male. Within a day, its ovaries shrink and its testes grow. If the same wrasse larva had reached a reef that had no males or that had territory undefended by a male, it would have developed into a male wrasse.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Conclusion

The extraordinary diversity of sex-determination mechanisms has long been noted. Many species resort to environmental cues for the determination of sex. A well-known example is temperature-dependent sex determination in many (although not all) reptiles. Yet, many other species use genetic mechanisms, i.e., sex chromosomes, for the determination of sex. It is generally accepted that environmental sex determination is the ancestral state and that genetic sex determination evolved as a derived condition. It is also recognized that genetic sex determination has originated on many independent occasions across diverse taxa.

Priyanka K. 8802387686

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 11: Sex Linkage

Introduction

Sex is chromosomally determined in many animals, such as humans and fruit flies, and in dioecious plants. In these cases, the **sex chromosomes** are the chromosomes essentially involved in sex determination and the **autosomes** carry the somatic function genes.

Sometimes, it is seen that the sex chromosomes also carry some genes for somatic characters apart from the primary sex controlling genes. In general, genes on sex chromosomes are called **sex linked genes** and the traits controlled by them is called **sex linked trait**. Thus, **sex-linked inheritance** is the transmission of somatic characters and their determining genes along with sex determining genes on the sex chromosomes. Sex linked inheritance was discovered by TH Morgan (1910) when he studied the inheritance of red / white eye colour trait in *Drosophila*.

Basic types of sex linked traits

Cytogeneticists have divided the X and Y chromosomes of some species into homologous and nonhomologous regions. The latter are called *differential* regions (Figure 1).

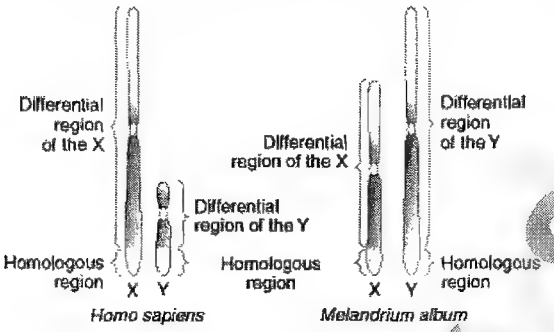


Figure 11: Differential and pairing regions of sex chromosomes of humans and of the plant *Melandrium album*. The regions were located by observing where the sex chromosomes paired up in meiosis and where they did not.

The differential regions contain genes that have no counterparts on the other sex chromosome. Accordingly, there are three basic types of sex-linked traits.

1. Genes in the differential region of the X show an inheritance pattern called **X linkage**. The X - chromosome which is common to male and female carries a number of genes. It is estimated that in *Drosophila melanogaster* the X- chromosome bears about 150 genes while in human beings it is 645 known genes (based on Macmillan Science Library; 2008 - *Genetics of X Chromosome*).

2. Genes in the differential region of the Y show **Y linkage**. Such genes are also called **Holandric genes.** These genes are passed as such to the male offspring. It is known as holandric inheritance, *e.g.,* TDF, hypertrichosis, porcupine skin. TDF is a very small but master gene which activates a series of genes required for male sex differentiation.

3. Genes in the homologous region show **X-and-Y linkage**. This is a rare condition, but present nevertheless, *e.g.,* bobbed or small bristles in *Drosophila*.

Inheritance pattern of sex linked traits

Sex linkage is of four types:

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

1. Diagenic (Diagynic). Criss-cross from a male parent to grandson through daughter of F1 generation.
2. Diandric. Criss cross from female parent to grand-daughter through son of F1 generation.
3. Hologenic. Directly from female to female or mother to daughter.
4. Holandric. Directly from male to male or father to son.

Except for direct inheritance (holandric, hologenic), the sex linked characters show crisscross inheritance.

Characteristics of Sex Linked Inheritance

1. It is criss-cross inheritance in most cases except direct inheritance (holandric, hologenic). Criss-Cross inheritance is a type of sex-linked inheritance where genes are transferred from a parent to the grand child of same sex through offspring of the other sex. It is of two types, diagenic and diandric.

   In the cross (shown in Figure 2) performed by Morgan (pure red eyed female x white eyed male in *Drosophila*), the white eye trait of the male fly is passed to male fly of F2 generation through female fly of F1 generation.

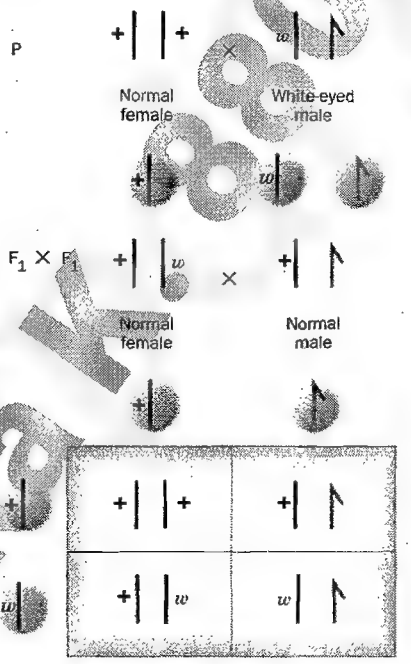


Figure 12: The cross performed by Morgan (pure red eyed female x white eyed male in *Drosophila*)

   The reason is that the male passes its Y-chromosome to the son and X-chromosome to the daughter. The daughter receives XX-chromosomes, one from father and the second from-mother. It passes its one X-chromosome to the male offspring and second to the female offspring.

2. Opposite behaviour is found in those cases where the males are homogametic while the females are heterogametic. The single recessive alleles find expression in females, as the

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

latter are hemizygous. A good example is the trait of bars on plumage in Plymoth Rock chicken (male ZZ, female ZW) 3. Homogametic parent *(e.g.,* XX mother in humans or ZZ father in poultry) passes sex-linked traits to both sons and daughters.

3. Majority of the sex-linked traits are recessive.

4. Heterogametic individuals *(e.g.,* human males) suffer from sex-linked disorders more often than the homogametic individuals *(e.g.,* human females) do.

5. Rate of mutations is higher in heterogametic individuals *(viz.,* human males).

6. Homogametic individuals *(e.g.,* human females) function as carriers of sex-linked disorders because recessive genes can express in them only in homozygous state.

7. In human beings (XX - XY sex complement) traits governed by sex-linked dominant genes *(e.g.,* defective enamel of teeth) produce disorders in females more often than in males. Sex Linked Traits. They are those traits the determining genes of which are found on the sex chromosomes. All the sex-linked traits present on a sex chromosome are inherited together.

Examples of sex linked traits

X-linked recessive traits

X-linked recessive traits are expressed in all heterogametics, but only in those homogametics that are homozygous for the recessive allele. In other words, the effects of this sex-linked, X chromosome disorder are manifested almost entirely in males, although the gene for the disorder is inherited from the mother.

Diseases well known for their X-linked recessive inheritance are hemophilia (types A and B), Duchenne muscular dystrophy and color blindness.

Haemophilia is a group of hereditary genetic disorders that impair the body's ability to control blood clotting or coagulation. In its most common form, Hemophilia A, clotting factor VIII is absent. In Haemophilia B, factor IX is deficient. Hemophilia A occurs in about 1 in 5,000–10,000 male births, while Hemophilia B occurs at about 1 in about 20,000–34,000.

Color blindness is a color vision deficiency in animals, involving the inability to perceive differences between some of the colors that others can distinguish. It is most often of genetic nature, but may also occur because of eye, nerve, or brain damage, or due to exposure to certain chemicals.

Genetically, the disease is inherited from recessive mutations on the X chromosome but the mapping of the human genome has shown there are many other causative mutations. In other words, mutations capable of causing color blindness originate from at least 19 different chromosomes and many different genes.

Duchenne muscular dystrophy (DMD) is a severe recessive x-linked form of muscular dystrophy that is characterized by rapid progression of muscle degeneration, eventually leading to loss in ambulation, paralysis, and death. This affliction affects one in 3500 males, making it the most prevalent of muscular dystrophies. In general, males are only afflicted, though females can be carriers. The disorder is caused by a mutation in the gene DMD, located in humans on the X chromosome. The DMD gene codes for the protein dystrophin, an important structural component within muscle tissue.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Y linked traits

Important Y linked genes are:

1. AMELY (amelogenin,Y-chromosomal)
2. ANT3Y (adenine nucleotide translocator-3 on the Y)
3. ASMTY (which stands for acetylserotonin methyltransferase)
4. AZF1 (azoospermia factor 1)
5. AZF2 (azoospermia factor 2)
6. BPY2 (basic protein on the Y chromosome)
7. CSF2RY (granulocyte-macrophage colony-stimulating factor receptor, alpha subunit on the Y chromosome)
8. DAZ1 (deleted in azoospermia)
9. DAZ2
10. IL3RAY (interleukin-3 receptor)
11. PRKY (protein kinase, Y-linked)

As the Y-chromosome is small and does not contain many genes, relatively few traits are Y-linked, and so Y-linked diseases are rare.

Priyanka K. 8802387667

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 12: Molecular basis of sex determination in plants

The genetic control of sex determination and differentiation is well understood in several animal systems, particularly *Drosophila melanogaster*, *Caenorhabditis elegans* and mammals.

In plants, the understanding of the sex determination and differentiation systems is not very clear – as most of them are *sexually monomorphic*, all individuals are essentially alike in their gender condition.

The term *cosexual* (Lloyd, 1984) is used when individual plants have both sex functions. There are two types of cosexual condition.

1. Present within each flower: Hermaphrodite
2. Present in separate male and female flowers but on the same plant: Monoecious.

A minority of plant species are *Sexually Polymorphic*, including dioecious species (Table 1).

Plant term	*Definition of plant term*	*Occurrence in plants, and examples*
Sexually monomorphic		
Hermaphrodite	Flowers have both male and female organs	90% of flowering plants (eg roses)
Monoecious	Separate sex flowers on the same individuals	5% of flowering plants, often those with catkins (eg hazel), and many gymnosperms (eg pines)
Gynomonoecious (male-sterility)	Individuals have both female and hermaphrodite flowers	eg daisies
Andromonoecious (female-sterility)	Individuals have both male and hermaphrodite flowers	
Sexually polymorphic		
Dioecious	Separate sex individuals (male and female plants)	5% of flowering plants (eg holly), and some gymnosperms
Gynodioecious	Individuals either female or hermaphrodite	eg ribwort plantain (*Plantago lanceolata*) bladder campion (*Silene vulgaris*)
Androdioecious	Individuals either male or hermaphrodite	very rare

Table 13: Sexual phenotypes in plants

Evolution of unisexuality and dioecy

Many dioecious plant species have hermaphrodite relatives. Some cases of dioecy have evident rudiments of opposite sex structures in flowers of plants of each sex. These instances suggest that the evolution of unisexual flowers is a recent phenomenon (Darwin, 1877) and hermaphroditism and monoecy are basal

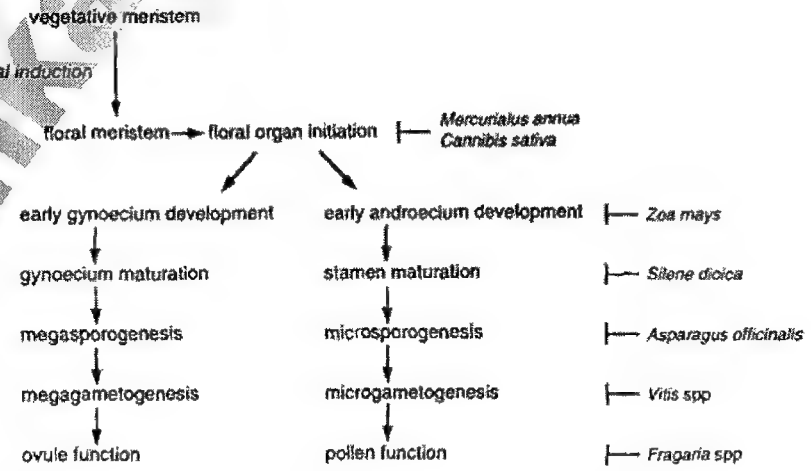


Figure 1: Developmental steps in floral development blocked during the sex determination process

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

conditions.

Many studies further confirm that the sex chromosomes in plants have evolved repeatedly and quite recently.

Molecular control of sex differentiation

1. The developmental stage of differentiation when sex is established

In different plant species, sex is established at different stages of floral development (Fig. 1). It means that the molecular pathways for sex differentiation operate differently in different taxa.

2. The studies on *Silene latifolia*

In *Silene latifolia*, the males are heterozygous with 2A:XY genotype. The female individuals are 2A:XX. The Y-chromosome carries three genes with male sex differentiation activity. The *Silene latifolia* Y chromosome, showing genes related to sex differentiation is illustrated below.

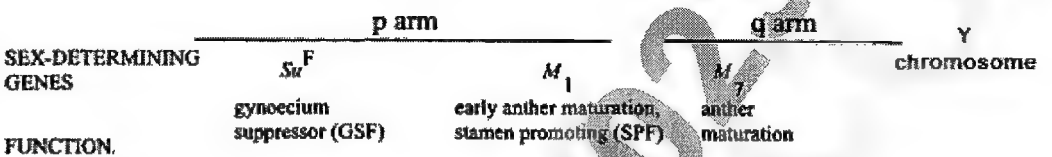


Figure 2: *Silene latifolia* Y chromosome

3. The studies in Asparagus

Asparagus officinalis shows dioecy. The critical stage for sex determination in *Asparagus* occurs much later in floral development. Flower buds from females and males are phenotypically same till the onset of meiosis.

At this time, pollen formation is arrested in female flowers and embryo sac formation is arrested in male flowers, so that the mature flowers are functionally unisexual.

The defect in stamen maturation in pistillate flowers is the precocious degeneration of the tapetal cells. It is controlled by suppression of the *Tm7* gene. Due to this there is collapse of the microspore mother cells.

In staminate flowers, degeneration begins in nucellar and integumentary cells and progresses to the megaspore mother cell (Lazarte and Palser, 1979). While the *Tm7* gene expresses normally. However, the gene responsible for degeneration in nucellar and integumentary cells are not known.

4. The role of Gibberellic Acids (GAs)

The plant hormones Gas have pronounced effect on male sexual development. GA stimulates the gene *GAMYB* of the *MYB* superfamily. MYB proteins are a superfamily of transcription factors that play regulatory roles in developmental processes and defense responses in plants.

GAMYB transcription factor stimulates the stamen formation and differentiation related genes in plants. Rice mutant lines lacking *GAMYB* action form male sterile flowers. In *Arabidopsis*, the mutations in genes *AtMYB33* and *AtMYB65* cause male sterily (Millar and Gubler, 2005).

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

5. Sex reversal in Papaya

Carica papaya has three sex chromosomes: the X, Y, and Y^h. XX produces a female plant, XY a male, and XY^h a hermaphrodite. All combinations of Y and Y^h fail to develop beyond the early embryonic stage after pollination.

The Y and Y^h chromosomes contain gene (*Sex1*) that promotes the development of the male reproductive organ, the stamen, in male and hermaphrodite trees (Ming, 2008). The Y chromosome also contains a gene that disables the development of the female sexual organ, the carpel.

Ming (2009) has shown that the Y^h chromosome lacks the gene that turns off development of the carpel, therefore allowing both male and female organs to grow in XY^h plants.

Male plants of *Carica papaya* can be induced to bear female flowers and yield a good fruit crop by the application of *ethrel* and *chlorflurenol* at various concentrations. This is known as *Sex Reversal*. The molecular pathway of sex reversal has been studied by J. Deputy (2002). He showed that sex reversal is caused by increased levels of C_2H_4.

It increases the expression of a feminizing or stamen suppressor gene (Sex 4), which causes stamen abortion before or at flower inception.

Priyanka K. 8802381681

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 13: Cytoplasmic inheritance and cytoplasmic genes

Introduction to Cytoplasm Inheritance

Cytoplasmic Inheritance is the transmission of genes that occur outside the nucleus. It is an **extra-nuclear** and **extra-chromosomal** a type of inheritance of traits via genes contained in cytoplasmic organelles viz. mitochondria and chloroplasts or from cellular parasites like viruses or bacteria.

Cytoplasmic inheritance was discovered by Correns (1908) in a variegated variety of the four-o'clock plant *Mirabilis jalapa*. In his experiment, Correns noted that the pattern Variegated plants have some branches, which carry normal green leaves, some branches with variegated leaves (mosaic of green and white patches), and some branches which have all white leaves. He also observed that inheritance of these features could not explained by Mendellian Mechanisms.

Similar observations were also made in the ornamental plant *Pelargonium zonale* by Baur in 1911.

Characteristic Features

The characteristic features of this type of inheritance include the following.

1. No equivalent result in reciprocal crosses. At most, the results can be nearly comparable.
2. In higher eukaryotes, the transmission pattern is mostly maternal. However, depending upon the fertilization system, the other type of uniparental, i.e., paternal inheritance may also be seen. For example in annelids, or even biparental transmission pattern can be seen.
3. The Mendellian segregation ratios are always absent.
4. There is an irregularity or patchiness in the genetic transmission of the trait showing no regular pattern. For this reason, there is no predictability about uniform phenotype in all the parts of the same individual.
5. These characters show no linkage to any of the known nuclear markers and experimental enucleation would have no effect on the expression of the trait.
6. Elimination of the cytoplasm will drastically affect the pattern of phenotypic expression of these traits.
7. The genetic transmission pattern is greatly affected by the pattern of cytokinesis, especially when it is asymmetrical.
8. A mutation arising within one organellar DNA molecule generates a mixture of mutant and wild-type DNA sequences within that cell. The occurrence of two distinct varieties of DNA within the cytoplasm of a single cell is termed *heteroplasmy*. When a heteroplasmic cell divides, the organelles segregate randomly into the two progeny cells in a process called *replicative segregation*, and chance determines the proportion of mutant organelles in each cell. Although most progeny cells will inherit a mixture of mutant and normal organelles, just by chance some cells may receive organelles with only mutant or only wild-type sequences; this situation is known as *homoplasmy*.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

9. Cytoplasmically controlled traits are normally of peripheral importance, rather than being directly related to survival or reproductive abilities of an organism.

10. In some cases, cytoplasmic factor mutations may lead to abnormal or diseased condition such as –

    - Cytoplasmic Male Sterility in higher plants
    - Leaf Variegation arising from the loss of plastidial genome segments
    - Poky condition in *Neurospora*
    - Petite condition in *Saccharomyces*
    - Leber's Hereditary Optic Neuropathy (LHON) in humans
    - Pearson's Marrow Pancreas Syndrome in humans

Classes of Extra-nuclear Inheritance

There are three different classes at extra-nuclear inheritance or non mendelian inheritance, *viz.*

1. Maternal effects
2. Inheritance due to infectious particles
3. Organellar genome inheriatance or "True" Cytoplasmic Inheritance

These are briefly described below with examples.

Maternal Effects

When the expression of a character is influenced by the nuclear genotype of female parent, it is referred to as maternal effect. Such characters exhibit clear-cut differences in F_1 for reciprocal crosses. Maternal effects are known both in plants and animals. The maternal effect genes normally express during the maturation of the egg cell. The proteins encoded by these gene stay in the egg cell prior to fertilization and then continue till the early stages of embryogenesis and exert profound influence on the pattern of embryonic development. Such effects seem to be independent of the contributions of the male parent, therefore known as maternal effects. Some examples of maternal effects include Coiling Pattern of Shell in Snail and Pigment in Flour Moth, *Ephestia kuhniella.*

Inheritance Involving Infective Particles

In some cases, cytoplasmic inheritance is associated with infective particles like parasite, symbiont or viruses, which are present in the cytoplasm of an organism. However, such cases are not considered as true examples of cytoplasmic inheritance. Some examples of this type are Kappa Particles in *Paramecium* and Sigma Particle in *Drosophila.* In *Paramecium* the Kappa Particle is due to the presence of a bacterium called *Cedobacter teniospiralis.*

Organellar Inheritance or the True Cytoplasmic Inheritance

The true cytoplasmic inheritance is one which involves the genomes of plastids (chloraplasts) and mitochondria. Thus, true cytoplasmic inheritance is again of two types, *viz,*

a. Plastid inheritance

b. Mitochondrial inheritance.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

The former is applicable to plants only because plastids are found only in plants. The mitochondrial inheritance is common for both plants and animals.

A brief description of plastidial and mitochondrial genome and inheritance is given below.

Mitochondrial Genome and Inheritance

In animals and most fungi, the mitochondrial genome consists of a single, highly coiled, circular DNA molecule Plant mitochondrial genomes often exist as a complex collection of multiple circular DNA molecules. Each mitochondrion contains multiple copies of the mitochondrial genome, and a cell may contain many mitochondria. A typical rat liver cell, for example, has from 5 to 10 mtDNA molecules in each of about 1000 mitochondria; so each cell possesses from 5000 to 10,000 copies of the mitochondrial genome, and mtDNA constitutes about 1% of the total cellular DNA in a rat liver cell. Like eubacterial chromosomes, mtDNA lacks the histone proteins normally associated with eukaryotic nuclear DNA. The guanine — cytosine (GC) content of mtDNA is often sufficiently different from that of nuclear DNA that mtDNA can be separated from nuclear DNA by density gradient centrifugation.

Mitochondrial genomes are small compared with nuclear genomes and vary greatly in size among different organisms.

Human mtDNA is a circular molecule encompassing 16,569 bp that encode two rRNAs, 22 tRNAs, and 13 proteins. The two nucleotide strands of the molecule differ in their base composition: the heavy (H) strand has more guanine nucleotides, whereas the light (L) strand has more cytosine nucleotides.

The organization of yeast mtDNA is quite different from that of human mtDNA. Although the yeast mitochondrial genome with 78,000 bp is nearly five times as large, it encodes only six additional genes, for a total of 2 rRNAs, 25 tRNAs, and 16 polypeptides. Most of the extra DNA in the yeast mitochondrial genome consists of noncoding sequences. Yeast mitochondrial genes are separated by long intergenic spacer regions that have no known functions.

Flowering plants (angiosperms) have the largest and most complex mitochondrial genomes known; their mitochondrial genomes range in size from 186,000 bp in white mustard to 2,400,000 bp in muskmelon. Even closely related plant species may differ greatly in the sizes of their mtDNA.

Part of the extensive size variation in the mtDNA of flowering plants can be explained by the presence of large direct repeats, which constitute large parts of the mitochondrial genome. Crossing over between these repeats can generate multiple circular chromosomes of different sizes. The mitochondrial genome in turnip, for example, consists of a "master circle" consisting of 218,000 bp that has direct repeats. Homologous recombination between the repeats can generate two smaller circles of 135,000 bp and 83,000 bp.

Exceptions to the universal genetic code exist in mitochondria, and these exceptions often differ among organisms. For example, AGA specifies arginine in the universal code, but AGA codes for serine in Drosophila mtDNA and is a stop codon in mammalian mtDNA.

The Mt DNA mutation accounts for conditions like:

- Cytoplasmic Male Sterility in higher plants
- Petite condition in *Saccharomyces*

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

- Poky condition in *Neurospora*
- Leber's Hereditary Optic Neuropathy (LHON) in humans
- Pearson's Marrow Pancreas Syndrome in humans

Plastisial Genome and Inheritance

Chloroplasts are the important plastids. Plastids have green pigments called chloroplasts. Plastids self duplicate, have some amount of DNA and play an important role cytoplasmic inheritance.

Most of our current knowledge about the chloroplast genome is based on the work of Ohyama, K., et al [1986]; Linda A. Raubeson and R. K. Jansen [1992] and Frank Walter [2001].

The genome of the chloroplasts found in *Marchantia polymorpha*, a liverwort contains 121,024 base pairs in a closed circle. These make up some 128 genes, which include:

- Duplicate genes encoding each of the four subunits (23S, 16S, 4.5S, and 5S) of the ribosomal RNA (rRNA) used by the chloroplast.
- 37 genes encoding all the transfer RNA (tRNA) molecules used for translation within the chloroplast.
- 4 genes encoding some of the subunits of the RNA polymerase used for transcription within the chloroplast.
- A gene encoding the large subunit of the enzyme RUBISCO (Ribulose bisphosphate carboxylase/oxygenase)
- 9 genes for components of Photosystems I and II
- 6 genes encoding parts of the chloroplast ATP synthase
- Genes for 19 of the ~56 proteins used to construct the chloroplast ribosome

All these gene products are used within the chloroplast, but all the chloroplast structures also depend on proteins encoded by nuclear genes, translated in the cytosol, and imported into the chloroplast. RUBISCO, for example, the enzyme that adds CO_2 to Ribulose bisphosphate to start the Calvin cycle, consists of multiple copies of two subunits:

a. A large one encoded in the chloroplast genome and synthesized within the chloroplast, and

b. A small subunit encoded in the nuclear genome and synthesized by ribosomes in the cytosol. The nuclear genes coding for the small sub-unit transcribe in a light dependent manner. The small subunit synthesized in the cytoplasm must be imported into the chloroplast.

Some examples of plastid inheritance include leaf variegation in *Mirabilis jalapa* and *Pelargonium zonale;* Iojap in Maize and streptomycin resistance in *Chlamydomonas.*

Significance of Cytoplasmic Inheritance

1. Cytoplasmic inheritance has been useful in explaining the role of various cytoplasmic organelles in the transmission of characters in different organisms.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

2. Studies of cytoplasmic inheritance have played key role in mapping of chloroplast and mitochondrial genome in several species.

3. Development of cytoplasmic male sterility or CMS in several crops like maize, pearl millet, Sorghum, cotton, etc. has taken place. Availability of CMS lines has facilitated the production of hybrid seed in these crops at a cheaper cost than with hand emasculation and pollination method. Role of mitochondria in the manifestation of heterosis is gaining increasing importance these days.

4. Mutation of chloroplast DNA and mitochandrial DNA leads to generation of new variants. Some of such variants arts of special significance especially in ornamental plants.

Priyanka K. 8802387[illegible]

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 14: Genetics of male sterility

Introduction to male sterility

Male sterility is a reproductive anomaly where an abnormality at some stage of male sexual development leads to failure of reproduction. In flowering plants the phenomenon was first reported in 1763 by Koelreuter.

In common parlance, male sterility in plants implies their inability to produce or release functional pollen and is the result of failure of formation or development of functional stamens, microspores or gametes.

Although the phenomenon is commonly described as pollen failure, the term is much broader in its scope. The male sterility may exist as pollen sterility i.e. non-functional pollen is formed; staminal sterility in which stamens are malformed or even absent; and functional sterility where viable pollen is trapped in indehiscent anthers. There can be other types of male reproductive failure, as enumerated below. Of the types discussed below, the pollen sterility is of most common type as a result of which the flowers generally do not produce functional anthers or viable pollen but the ovaries function normally.

Male sterility can be an outcome of any of the following abnormality in angiosperms.

1. Failure of microspore mother cell to undergo normal meiosis.
2. Failure of newly formed microspore to undergo maturation process. It often occurs due to abnormality of tapetum.
3. The indehiscent nature of anther which does not allows pollens to escape within the period of viability.
4. Inadequate or no deposition of Pollenkitt substances in the Pollen wall of entomophilous pollens, because of which pollen dispersal through insect vectors does not occur.
5. Failure of the Pollen to germinate on a compatible stigma.
6. Inadequate growth of the pollen tube in the stylar canal.
7. An abnormal gamete formation within the pollen tube. Such gametes fail to fuse with the egg.

As a result of male sterility, the flowers cannot be self-pollinated but they can be cross-pollinated. This attribute makes the male sterile system useful to the plant breeder especially for the production of hybrid seed. A recent account of history and classification of male sterility is given by Kaul (1988, 1998).

Characteristics & Types of Male Sterility

Characteristics

1. It may be an outcome of any abnormality in a broad spectrum of events leading to the formation of Male gamete.
2. It is an important outbreeding device that prevents autogamy and promotes allogamy.
3. It leads to heterozygosity due to elimination of inbreeding.
4. It is mostly genetic in control; however chemical substances and environmental changes can also induce Male sterility.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

5. Genes governing male sterility may be nucleus located or mitochondria located or both.
6. Transgenic technology can also induce male sterility.
7. It is found in a large number of angiosperms in the wild (approximately 25%).
8. This Phenomenon is widely applied in hybrid production. (Basis: As a result of male sterility, the flowers cannot be self-pollinated but they can be cross-pollinated. This attribute makes the male sterile system useful to the plant breeder especially for the production of hybrid seed).

Types

A recent account of history and classification of male sterility is given by Kaul (1988, 1998). Male sterility is most conveniently classified based upon the underlying causes. Accordingly, the phenomenon of male sterility can be grouped as follows:

Genetically Controlled Male Sterility

1) Nuclear male sterility (NMS): Failure of pollen formation is because of one or more nuclear (dominant or recessive) genes.

2) Cytoplasmic male sterility: Commonly known as CMS, it is based on the cytoplasmic (mitochondria located) genes only.

3) Cytoplasmic-genetic male sterility: Commonly known as CGMS, in which male sterility is due to the interaction between nuclear and cytoplasmic (mitochondrial) genes.

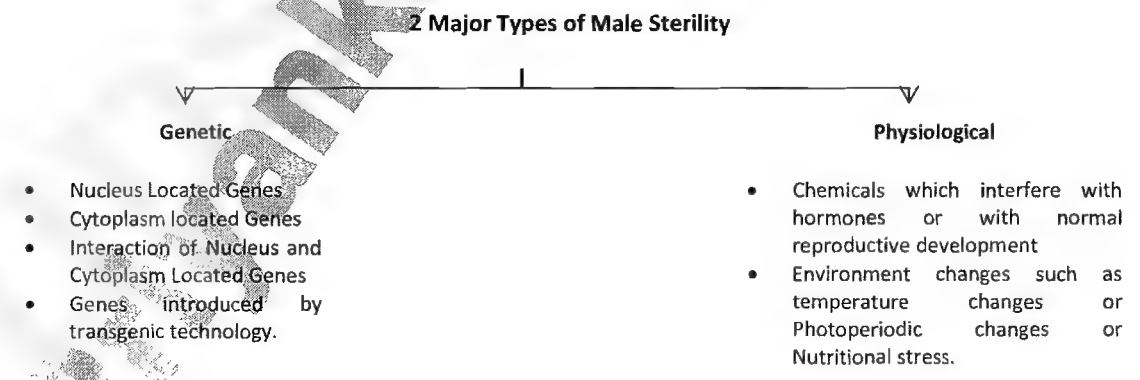
4) Transgenically caused male sterility: This type of male sterility arises due to the introduction of some gene into the plant from some other organism by recombinant DNA technology.

Physiologically Controlled Male Sterility

1) Gene-environment induced male sterility: Male sterility is induced due to interaction between nuclear genes and environmental conditions.

2) Chemically caused male sterility: Here, the application of certain chemicals on the plant at a particular stage of the reproductive development process causes male sterility.

The following chart summarizes the currently accepted classification of male sterility.



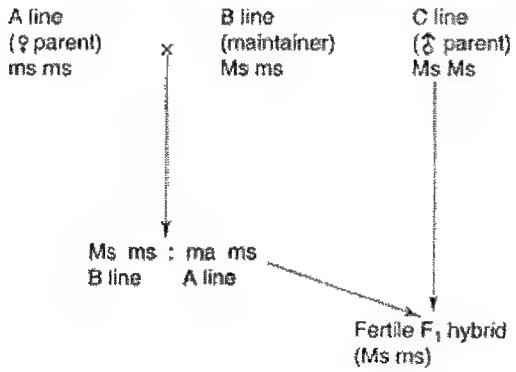
Detailed discussion of individual types

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Nuclear Male Sterility (NMS)

Nuclear male sterility also called as 'genetic or genic' male sterility (GMS), is usually caused by a recessive allele 'ms'. Nuclear male sterility originating through spontaneous mutation is a common phenomenon in nature. It is of wide occurrence in flowering plants and as many as 60 genes for male sterility in maize, 55 in tomato, 10 in cotton and 60 in rice are known (Homer and Palmer, 1995). The dominant allele 'Ms' results in the development of normal anthers and pollen. Therefore, 'ms ms' genotype is male sterile, while 'Ms ms' or 'Ms Ms' are male fertile. It has not been largely used for hybrid seed production because it is not possible to produce a population of 100 per cent male sterile plants by this method as the maximum percentage of male sterile plants that can be produced is 50%. The completely male sterile plants (ms ms) are crossed with heterozygous (Ms ms) to produce a progeny of Msms and msms out of which 50 % are fertile and that can be crossed with steriles to maintain male sterility. The nuclear male sterility may originate through spontaneous mutation or can be induced through physical or chemical mutagenesis.



Nuclear male sterility and hybrid seed production

Nuclear male sterility is usually recessive and monogenic as a result of which fertility restoration in the hybrid and crossing scheme are relatively easy. In this case, roguing of fertile heterozygous (Ms ms) plants is essential in the seed production plots. As shown in figure here, pure breeding male sterile lines cannot be maintained unless fertility is restored by a modified environment as in case of photosensitive genetic male sterility system (PGMS) where same line can be made male sterile or fertile under different photoperiodic conditions. Maintenance of male sterile (A) line needs identification of heterozygous fertile plants (B) which are used to pollinate male sterile plants to produce seed that shall segregate for male sterility and fertility in the ratio of 50:50. For hybrid seed production the fertile plants from male sterile line are to be identified and rogued before anthesis. This could be achieved by morphological marker genes closely linked to male sterility locus. Several characters linked with 'ms' gene have been reported as white endosperm in maize, shrunken endosperm in barley, short leaf in sorghum and woolly character in tomato. However, insufficient linkage would result in recombination between marker and male sterility gene as a result of which the marker aided identification becomes invalid. The male sterility conditioned by dominant genes has been reported in cotton, wheat, carrot and oil-seed rape but it cannot be used in hybrid seed production as the F1 hybrid is rendered male sterile. The commercial hybrid seed is produced by crossing the male sterile plants with another line C which is' homozygous for fertility. If seed is the economic part then we have to be sure that recessive male sterility gene (ms) is not present in the C line. Thus the production of hybrid seeds utilizing nuclear male sterility is more complicated, cost and labour intensive.

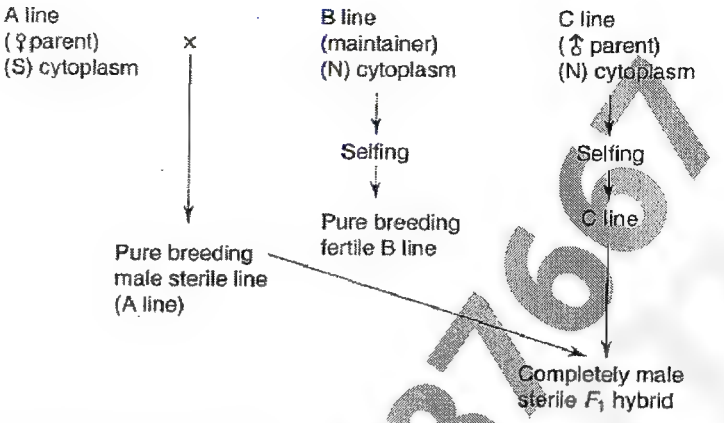
Cytoplasmic Male Sterility

The male sterility governed solely by cytoplasmic factors without any dependence on nuclear genes is termed as cytoplasmic male sterility. The Cytoplasmic genes responsible for male sterility are located in the mitochondria and not in the chloroplasts.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Since the offsprings inherit the cytoplasm of only the female parent, so the cross of cytoplasmic male sterile plants leads to completely male sterile progeny. Though cytoplasmic male sterility is a widely used term but such a class may only be reflecting the lack of suitable genes which actually interact with cytoplasm to regulate sterility and fertility. Kaul (1998) expressed that male sterility controlled exclusively by cytoplasmic genes i.e. cytoplasmic male sterility does not exist and is a total misnomer. The true cytoplasmic male sterility should remain uninfluenced by nuclear genes but such stable male sterile types are unknown as fertility restorer genes and maintainer nuclear genes have been detected in many plants (Kaul, 1988). All cytoplasmic male sterility may only be a form of genic-cytoplasmic male sterility which is so labelled till suitable interacting nuclear genes are identified. It may not be improbable that all presently known cytoplasmicgenic male sterility was at one time only cytoplasmic male sterility.

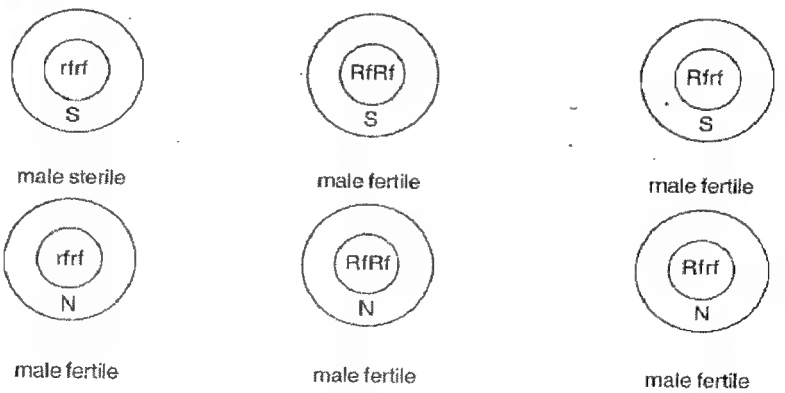


Cytoplasmic Male Sterility and Hybrid Seed Production

In case of cytoplasmic male sterility, the nuclear genes do not play any role and it is based solely on cytoplasmic (mitochondrial) genes transmitted maternally. The hybrids being governed by cytoplasm of female parent are thus always male sterile. Therefore, this system is useful only in crops where seed production is not important. Maintenance of the parental lines and production of hybrid seeds through the use of cytoplasmic male sterility is shown in Figure above.

Cytoplasmic-Genetic Male Sterility

Cytoplasmic-genetic male sterility results from the interaction among the cytoplasmic (mitochondrial) and nuclear genes. Cytoplasmic genetic male sterility (CGMS) and cytoplasmic male sterility (CMS) are used interchangeably. It has been reported in several hundred species (Kaul, 1988) and is being widely used for production of hybrid seed both in self- and cross-pollinated crops (sorghum, sunflower, pearl millet, maize, rice and oilseed rape). In this system, a specific mutation in the mitochondrial DNA

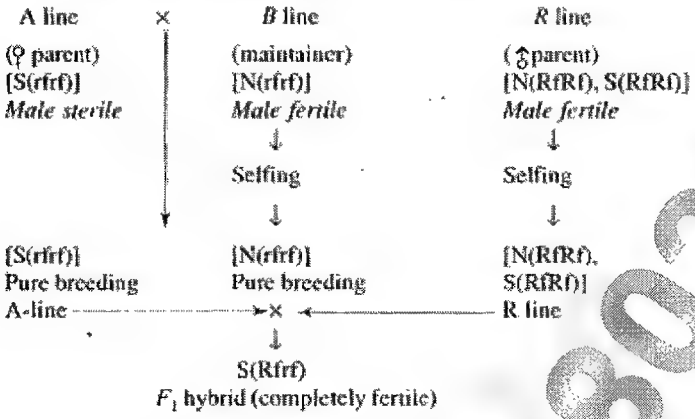


Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

(mtDNA) in combination with proper nuclear background leads to the failure of mature pollen formation. In some cases, the nuclear genes often called fertility restoration genes can compensate for the cytoplasmic mutation and normal pollen formation occurs. Therefore the expression of mitochondrial (cytoplasmic) male sterility. genes is governed by the type (dominant/recessive) of nuclear fertility restoring genes. The fertility restoring alleles have been represented by symbols like Rf (fertility restoring) in wheat, sunflower and corn, and Ms (male fertile) in onions, sorghum and pearl millet. Different combinations of sterile (S), normal (N) cytoplasms and fertility restoring genes determining the sterility/fertility of the plant are given in figure below.

Use of CGMS for commercial hybrid seed production



Normal (N) cytoplasm, irrespective of the gene in the nucleus results in a fertile plant. Sterile (S) cytoplasm operates only when both the alleles of fertility restoring nuclear gene are recessive (rfrf). Whereas, in the presence of dominant fertility restoring allele, the sterile cytoplasm becomes inoperative and the plant is normal and fertile. Generally the use of CGMS for commercial hybrid seed production involves the use of three breeding lines, namely:

1) **A-line:** CGMS-line or male sterile containing sterile cytoplasm and homozygous nuclear male sterility gene [S(rfrf)],

2) **B-line:** maintainer, cytoplasmic male fertile line carrying normal cytoplasm but nuclear gene for sterility [N(rfrf)] and is isogenic to A-line. It is used to multiply the CGMS line.

3) **R-line:** Restorer line, male parent, carries gene(s) [N(RfRf), S(RfRf)] which masks the expression of CGMS trait and restores fertility to the F1 hybrid.

A scheme for maintenance of parental lines and commercial hybrid seed production using cytoplasmic-genetic male sterility is shown here.

Cytoplasmic-genetic male sterility has already found large scale application in the production of hybrid seeds especially in crops like sorghum, sugar beet, corn, onion and carrot.

Male Sterility by Transgenesis

Recombinant DNA technology has been successfully used to produce male sterile lines. An engineered male sterility has not been used except for *barnase-barstar* system but this is likely to be more important in future hybrid-breeding programmes. The *barnase* gene is obtained from *Bacillus amyloliquefaciens*. The gene encodes an RNase, which degrades the tapetal RNA and prevents pollen maturation.

The restorer gene in this system is *barstar*.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Gene-Environment Induced Male Sterility

In some crops the expression of nuclear male sterility genes is governed by the prevailing environmental conditions. In such cases nuclear genes cause male sterility under specific environmental conditions and fertility is restored when the environmental conditions change. Thermo-sensitive genetic male sterility (TGMS) and photosensitive genetic male sterility (PGMS) are the two well studied gene-environment induced male sterility systems in rice (Virmani, 1994). These systems are called two line male sterility systems where there is no need of separate CMS and maintainer lines because male sterile line becomes fertile under specific environments that obviate the need of a maintainer line. But the major limitation of these systems is the availability of such genes and their instability under slight variable environmental conditions which are not predictable.

Chemical	*Sterility (Optimum treatment)*		*Remarks*
	Male	*Female*	
Dalapon	Partial	Severe	Severe plant damage
DPX 3778	High	Slight	Shorter plant, leaf burn, delayed flowering, toxic residue in seed
Estrone	Poor	Ovule damage	Plant damage
Ethephon (Ethrel)	High	Significant	Height and seed size reduced, late tillers fertile, dose cultivar specific
Gibberelic acid	Poor	Nil	Increased plant height
Hybrex	High	Slight	Chlorosis risk of damage
Maleic hydrazide	Partial	Severe	Severe plant damage and late maturity
Mendok (FW 450)	Poor	Significant	-do-

Chemical Control

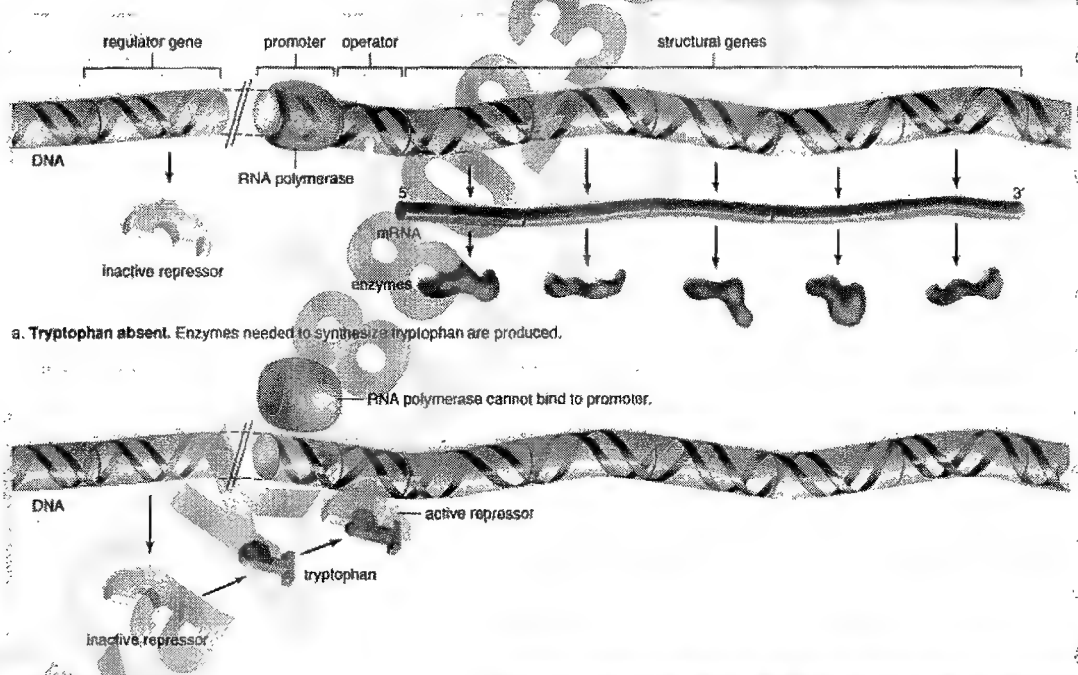
Induction of male sterility through application of chemicals was demonstrated as early as 1950 (Moore, 1950; Naylor, 1950). These chemicals known as male gametocides, male sterilants, pollenocides, androcides, pollen suppressants or more commonly as chemical hybridizing agents (CHAs) are applied to induce transitory male sterility. The effect of these chemicals may range from aberrant microspore meiosis to formation of defective pollen grains. For example, the application of ethrel to wheat anthers results in additional pollen mitosis and in barley ethrel causes degeneration of tapetal cells. This approach makes use of two line systems; any parent can be used as female, there is no need of maintenance of male sterile line and it eliminates the requirement of the fertility restoring gene. The expression of CHA-induced male sterility is highly stage specific and sensitive to environmental factors that leads to incomplete male sterility. Most of the CHAs have one or combination of undesirable effects on plant (Pickett, 1993). Some newly developed CHAs like 'Generis' by Monsanto and 'Croissor' by Rybrinova possess less undesirable effects on the plant. In general CHAs are applied as foliar spray prior to flowering that inhibits the formation of viable pollen but does not affect the seed development. Chemical hybridizing agents have been tried on cotton, corn, wheat, sorghum with varying degree of success. Several commercial hybrids of rice have been developed through the application of CHA in China.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

evolution
A Centre of Excellence for Civil Service Examination Guidance
B 11, Second Floor, Commercial Complex, Batra Cinema Road
Next to ICICI A.T.M., Dr. Mukherjee Nagar, DELHI - 110 009
Phone: 011 - 32974645, 32974651, 47092329

Genetics - II



a. **Tryptophan absent.** Enzymes needed to synthesize tryptophan are produced.

Priyanka Singh

© Copyright Evolution Educare Pvt. Ltd.

All rights are owned by or controlled by Evolution Educare Pvt. Ltd. The content of this book cannot be copied, reproduced, republished, posted, broadcast or transmitted in any way without first obtaining Evolution Educare Pvt. Ltd.'s written permission.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Detailed Table of Contents

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Priyanka 8802387667

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

evolution
Priyanka K. 8802387651

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Prescribed syllabus of Genetics

For Civil Service Examination

Covered in Genetics Book – I (Transmission Genetics)

Development of genetics; Gene versus allele concepts (Pseudoalleles); Quantitative genetics and multiple factors; Incomplete dominance, polygenic inheritance, multiple alleles; Linkage and crossing over; Methods of gene mapping, including molecular maps (idea of mapping function); Sex chromosomes and sex-linked inheritance, sex determination and molecular basis of sex differentiation; Cytoplasmic inheritance and cytoplasmic genes (including genetics of male sterility).

Covered in Genetics Book – II (Molecular Genetics)

Structure and synthesis of nucleic acids and proteins; Genetic code and regulation of gene expression; Gene silencing; Multigene families; Organic evolution – evidences, mechanism and theories. Role of RNA in origin and evolution.

Mutations (biochemical and molecular basis)

For Indian Forest Service Examination

Covered in Genetics Book – I (Transmission Genetics)

Development of genetics, and gene versus allele concepts (Pseudoalleles). Quantitative genetics and multiple factors. Linkage and crossing over–methods of gene mapping including molecular maps (idea of mapping function). Sex chromosomes and sexlinked inheritance, sex determination and molecular basis of sex differentiation. Cytoplasmic inheritance and cytoplasmic genes (including genetics of male sterility).

Covered in Genetics Book – II (Molecular Genetics)

Structure and synthesis of nucleic acids and proteins. Genetic code and regulation of gene expression. Multigene families. Organic evolution-evidences, mechanism and theories. Role of RNA in origin and evolution.

Mutation (biochemical and molecular basis).

Prions and prion hypothesis. (Already covered in the book on MICROBIOLOGY)

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 1: DNA Structure

Deoxyribonucleic acid (DNA) is a linear polymer of deoxyribonucleotides, which is found mostly in the form of a double helix and forms the universal genetic material in all life forms except for RNA viruses, Viroids and Virusoids. It is rarely in a single stranded form (as in some viruses like φX174).

Apart from role as genetic material, the DNA sequences are also involved in various other functions like regulation, structural stability and biocatalysis as DNAzymes.

The Structure of DNA

DNA has a multilevel structural organization, which can be grouped as:

1. Primary structure of DNA
2. Secondary structure of DNA
3. Tertiary structure of DNA

Primary Structure of DNA

Primary structure of DNA is a polymeric chain of deoxyribonucleotides. This chain is always linear in case of eukaryotic cells and many viruses, while in bacterial cells and in some viruses it is a covalently closed (circular) molecule. The repeating units of a DNA chain are nucleotides, each comprising three parts:

1. a pentose sugar- that is Deoxyribose,
2. a phosphate, and
3. a nitrogenous base.

1. The sugars of nucleic acids: It is called pentose sugars and has five carbon atoms. The sugars of DNA and RNA are slightly different in structure. RNA's ribose sugar has a hydroxyl group attached to the 2′-carbon atom, whereas DNA's sugar, called deoxyribose, has a hydrogen atom at this position. This difference gives rise to the names ribonucleic acid (RNA) and *deoxy*ribonucleic acid (DNA).

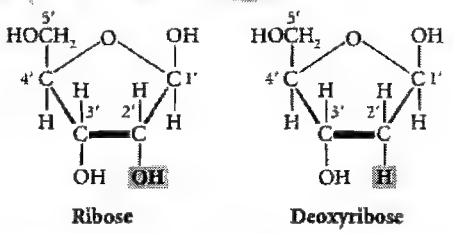


Figure 1: The sugars of nucleic acids

2. The phosphate group consists of a phosphorus atom bonded to four oxygen atoms. Phosphate groups are found in every nucleotide and frequently carry a negative charge, which makes DNA acidic. The phosphate group is always bonded to the 5′-carbon atom of the sugar in a nucleotide.

3. Nitrogen containing bases may be of two types: a purine or a pyrimidine.

A. **Purine**: Each purine consists of a six-sided ring attached to a five-sided ring. Both DNA and RNA contain two purines, adenine and guanine (A and G), which differ in the positions of their double bonds and in the groups attached to the six-sided ring.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

B. **Pyrimidine**: Each pyrimidine consists of a six-sided ring only. Three pyrimidines are found in nucleic acids: cytosine (C), thymine (T), and uracil (U). Cytosine is present in both DNA and RNA; however, thymine is restricted to DNA, and uracil is found only in RNA. The three pyrimidines differ in the groups or atoms attached to the carbon atoms of the ring and in the number of double bonds in the ring.

In a nucleotide, the nitrogenous base always forms a covalent bond with the 1′-carbon atom of the sugar. A deoxyribose (or ribose) sugar and a base together are termed as a nucleoside.

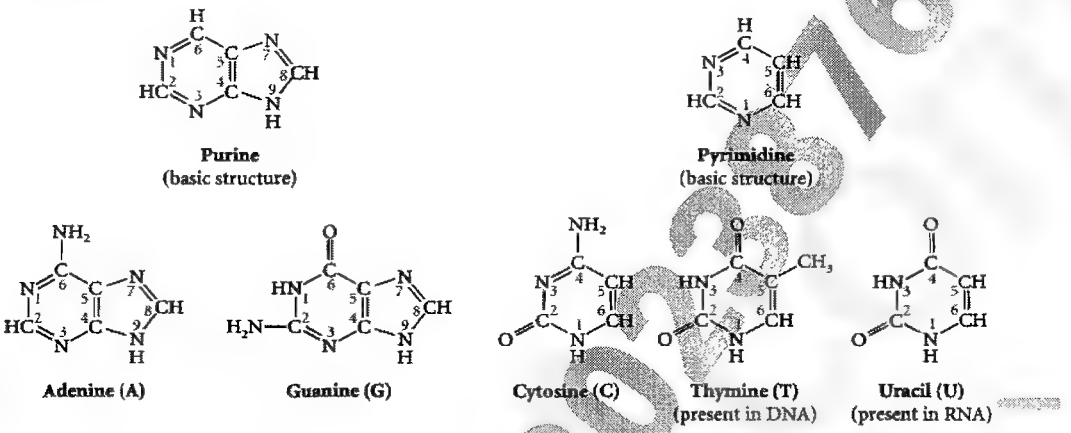


Figure 2: The nitrogenous bases present in nucleic acids.

The DNA nucleotides are properly known as deoxyribonucleotides or deoxyribonucleoside 5′-monophosphates. Because there are four types of bases, there are four different kinds of DNA nucleotides.

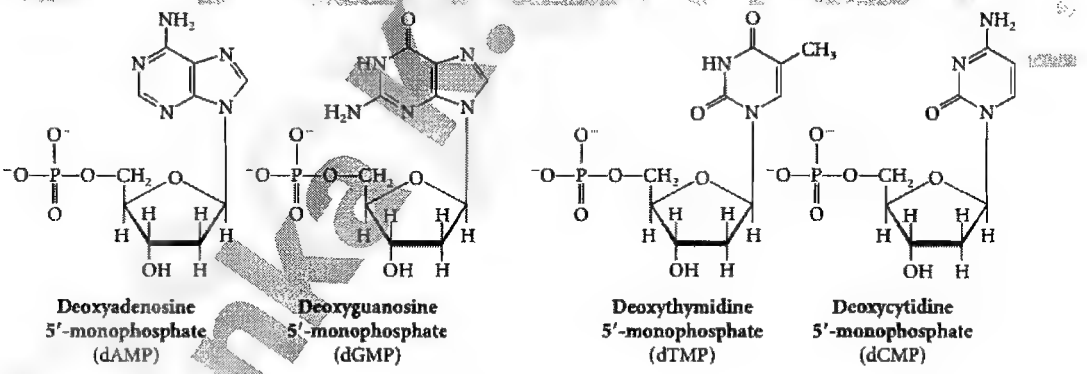


Figure 3: The four types of nucleotides found in DNA

Polynucleotide strands arise when many nucleotides are connected by covalent bonds called phosphodiester linkages, which join the 5′-phosphate group of one nucleotide to the 3′-carbon atom of the next nucleotide. The backbone of the polynucleotide strand is composed of alternating sugars and phosphates.

An important characteristic of the polynucleotide strand is its direction, or polarity. At one end of the strand, a phosphate group is attached to the 5′-carbon atom of the sugar in the nucleotide. This end of the strand is therefore referred to as the 5′ end. The other end of the strand, referred to as the 3′ end, has an OH group attached to the 3′-carbon atom of the sugar.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

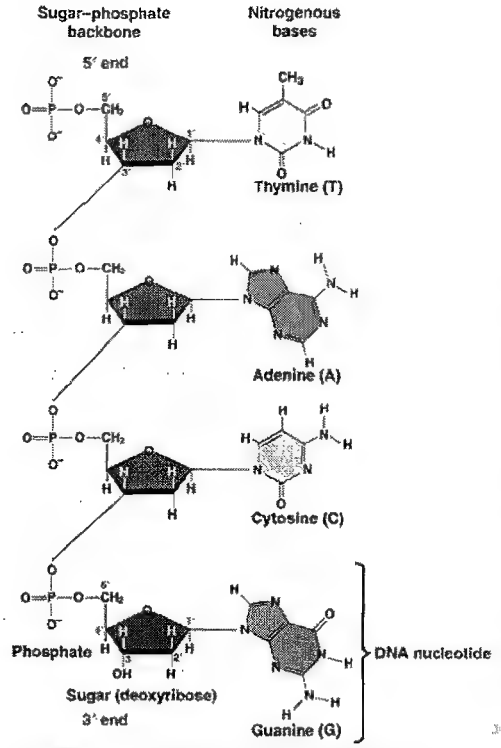


Figure 4: Nucleotide chain in DNA

Secondary structure of DNA

DNA is not a single molecule, but a pair of molecules joined complementarily by hydrogen bonds. The pair of polynucleotide chains describes the secondary structure of DNA.

A fundamental characteristic of DNA's secondary structure is that it consists of two polynucleotide strands wound around each other, forming a double helix. The sugar–phosphate linkages are on the outside of the helix and the bases are stacked in the interior of the molecule.

The details of this double helical structure were provided by James Watson and Francis Crick in 1953, when they published a paper in the British journal *Nature*. (J. D. Watson and F. H. C. Crick, "*Molecular Structure of Nucleic Acids: A Structure for Deoxynucleic Acids*." Nature 171 (1953): 738). The double stranded helical structure of DNA is also called the Watson and Crick Model.

The double helical structure of DNA is based on the following experimental findings of other scientists.

1. The base equivalence rule of Erwin Chargaff: Chargaff and co-workers discovered that in all organisms the total amount of adenine is always equal to the amount of thymine (A = T), and the amount of guanine is always equal to the amount of cytosine (G = C). In other words: A+G = T+C (Total purines = Total pyrimidines).
2. X-ray diffraction data of DNA from Maurice Wilkins and Rosalind Frank which showed that DNA is a right handed helical molecule.
3. Tautomeric data on the nitrogenous base from Jerry Donohue at the Cold Spring Harbour Lab showed that *in vivo* Adenine and Cytosine stay mostly in their Amino tautomeric form and Thymine and Guanine remain mostly in their Keto tautomeric form. This is an important determinant of base pairing.

The double helical DNA structural model of Watson and Crick says that:

1. The two polynucleotide chains in the double helix associate by *hydrogen bonding between the complementary nitrogenous bases.* In its *in vivo* tautomeric form, G can hydrogen bond specifically only with C, while A can bond specifically only with T.
2. Two hydrogen bonds stabilize the base pairing between A and T, while three hydrogen bonds stabilize the base pairing between G and C. Therefore, a GC base pair is difficult to separate.
3. The two polynucleotide chains to run in opposite directions (antiparallel).
4. The bases lie on the inside. They are flat structures, lying in complementary pairs perpendicular to the axis of the helix.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

5. Adjacent bases in a single strand are separated by 3.4 Å. The helical structure repeats every 34 Å, so there are 10 bases (= 34 Å per repeat/3.4 Å per base) per turn of helix. *However, a number of recent studies indicate that in vivo there are 10.4 nucleotides per complete turns, at least in eukaryotic cells.*

6. Each base pair is rotated ~36° around the axis of the helix relative to the next base pair. So ~10 base pairs make a complete turn of 360°.

7. The twisting of the two strands around one another forms a double helix with a minor groove (~12 Å across) and a major groove (~22 Å across)

8. The diameter of the helix is 20 Å.

9. Apart from Hydrogen bonds between complementary bases, the second force that holds the two DNA strands together is the interaction between the stacked base pairs. These stacking interactions contribute to the stability of the DNA molecule and do not require that any particular base follow another.

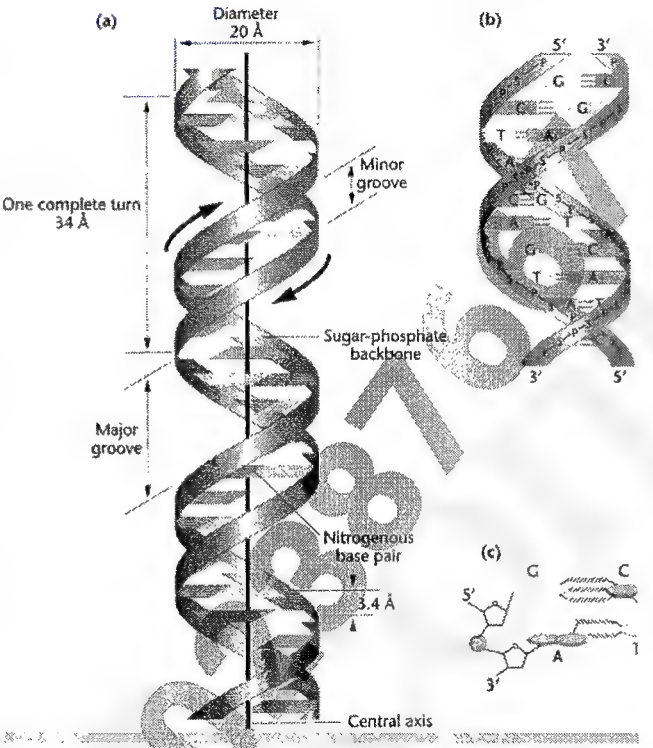


Figure 5: a. The double helical model of DNA. B. Base pairing C. Planar nature of base pairing

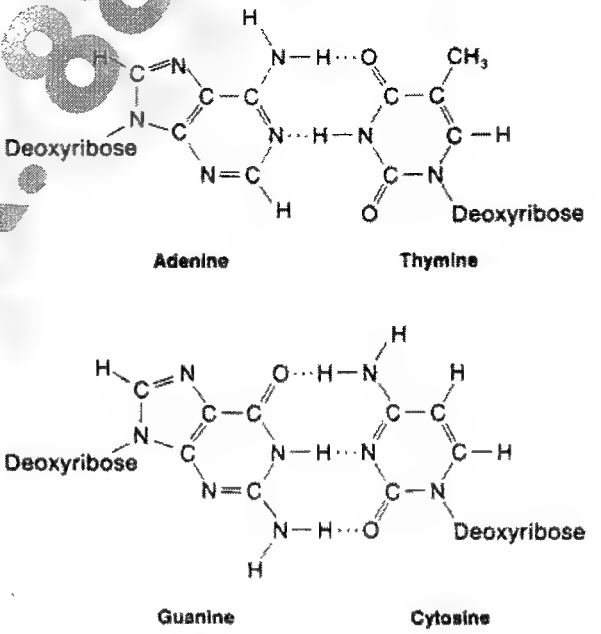


Figure 6: Base Pairing in DNA

Tertiary structure of DNA

In all organisms DNA is complexed with structural proteins to give rise to a condense and stable structure of a chromosome. A chromosome is thus a nucleoprotein complex that is the physical unit of genetic transmission. This is the tertiary structure of DNA.

Different Forms of DNA

DNA normally consists of two polynucleotide strands that are antiparallel and complementary.

The precise three-dimensional shape of the molecule can vary depending on the conditions in which the DNA is placed and, in some cases, on the base sequence itself.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Characteristic	A-DNA	B-DNA	Z-DNA
Conditions required to produce	75% H_2O	92% H_2O	Alternating purine structure and pyrimidine bases
Helix direction	Right-handed	Right-handed	Left-handed
Average base pairs per turn	11	10	12
Rotation per base pair	32.7°	36°	–30°
Distance between adjacent bases	0.26 nm	0.34 nm	0.37 nm
Diameter	2.3 nm	1.9 nm	1.8 nm
Overall shape	Short and wide	Long and narrow	Elongated and narrow

Significance of DNA Structure

1. The double-helical model of DNA and the presence of specific base pairs immediately suggested how the genetic material might replicate. As DNA is replicated, one of the chains of each daughter DNA molecule would be newly synthesized, whereas the other would be passed unchanged from the parent DNA molecule. This distribution of parental atoms is achieved by *semi conservative replication.*

2. The genotype which governs the organism's phenotype resides within the base pair sequences of DNA. Any alteration in this base sequence causes a change of genetic information or mutation.

3. The three major pathways of information transfer within the cell are DNA replication transcription, and translation. All these processes depend on DNA structure.

4. In molecular genetic studies; the DNA structure is a source of insight into key genetic processes such as Replication, Transcription, Recombination, Mutation and Transposition.

5. To maintain the fidelity of genetic transmission only, the cell maintains an elaborate set of DNA Repair and Proof Reading enzymes.

Priyanka 9538763

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 2: Transcription

The genomic control of phenotype & the central dogma

the phenotype of any organism is the outcome of the coordinated metabolic activities, which go on in the organism. Metabolism occurs by sequences of enzyme-catalyzed reactions, with each enzyme specified by one or more genes. Therefore, each step in a metabolic pathway and ultimately the organism's phenotype is under genetic control. To summarize, the flow of control is as follows.

Genes → Enzymes → Metabolic processes → Phenotypic expression

The flow of genetic information follows the central dogma of molecular biology in all the organisms of the living world. The *Retroviruses are the only known exceptions to the central dogma.* The central dogma of gene expression can be simplified by the following flow diagram.

Genetic Information in DNA → Transcription → **mRNA** (the intermediate message conveyors) → Translation → **Proteins**

Transcription

Of the two broad steps of gene expression, viz. transcription and translation, transcription is the first major process. In this, RNA is synthesized by using a specified DNA segment as template. The RNA molecule thus produced is called a **transcript**. All cellular RNAs are synthesized from a DNA template through the process of transcription.

Transcription is the universal first step in the transfer of genetic information from genotype to phenotype and biochemically it is the synthesis of RNA molecules using RNA polymerase as the key enzyme and with DNA as a template. The synthesis of an RNA molecule from DNA is a complex process involving one of the group of RNA polymerase enzymes and a number of associated proteins. The nucleotide [nt] sequence of the transcript is always complementary to the template strand of DNA.

The transcription unit is shown in Figure1. A **transcription unit** is a stretch of DNA that codes for an RNA molecule and the sequences necessary for its transcription. Within a transcription unit are three critical regions: a Promoter, an RNA-coding Region, and a Transcription Termination Site.

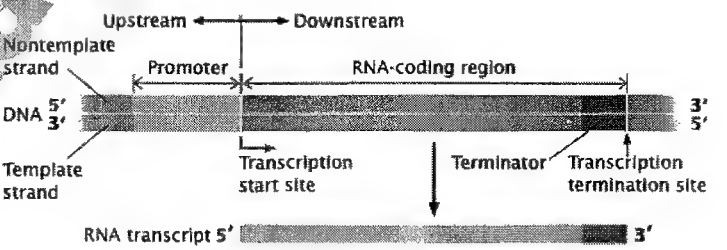


Figure 1: A Transcription Unit

1. The **promoter** is a DNA sequence that the transcription apparatus recognizes and binds. Transcription apparatus comprises of RNA polymerase and certain proteins (known as Transcription Factors) which help transcription. The promoter indicates which of the two DNA strands is to be read as the template and the direction of transcription. It also determines the transcription start site, the first nucleotide that will be transcribed into RNA. In most transcription units, the promoter is located next to the transcription start site but is not, itself, transcribed.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

2. The second critical region of the transcription unit is the **RNA-coding region**, a sequence of DNA nucleotides that is copied into an RNA molecule.

3. The third component of the transcription unit is the **Transcription Termination Site**, a sequence of nucleotides that signals where transcription is to end. These sites are also called Terminators. The Terminators are well defined in prokaryotic genomes but not so in Eukaryotes.

GENERAL FEATURES of TRANSCRIPTION

1. Broadly, the process appears to be similar to DNA replication, where one DNA strand acts as a template for the **5'→3' directed synthesis** of a complementary strand. However, a profound difference lies in the chemical nature of the complementary molecule. In transcription, it is RNA. Hence, instead of deoxyribonucleotides, ribonucleotides are found. Moreover, Thymine is replaced by Uracil.

2. During Transcription in any given region, only one DNA strand is used as template, while in replication both the strands of parent DNA molecule act as template.

3. RNA chain synthesis does not require any primer; it can start *de novo*.

4. The reaction is catalyzed by enzymes called **RNA polymerases.**

5. For the RNA synthesis, the polymerases need ribonucleotidetriphosphates [rNTPs]. This is a similarity with DNA polymerases.

6. RNA Polymerases can initiate transcription only at specific nucleotide sequences. Such nucleotide sequences are designated as **promoter sequences.** Therefore, promoter sites are the locations on the DNA molecule where transcription begins.

7. RNA polymerases are also capable of opening the DNA double helix locally. Hence, there is no need of helicases. Moreover, as transcription proceeds the same polymerase reseals the double strand too. By this mechanism, generation of torsion is avoided. The locally unwound site of DNA molecule is called a **transcription bubble**.

8. The process of transcription can also be subdivided into three stages:

    A. Initiation

    B. Elongation

    C. Termination

9. Molecular biologists often use the terms *upstream* and *downstream* to refer to the direction of transcription and the location of nucleotide sequences surrounding the RNA-coding sequence. The transcription apparatus is said to move downstream during transcription: it binds to the promoter (which is usually upstream of the start site) and moves toward the terminator (which is downstream of the start site).

Transcription in prokaryotes

Transcription in Bacteria can be conveniently divided into three stages:

1. initiation, in which the transcription apparatus assembles on the promoter and begins the synthesis of RNA;

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

2. elongation, in which RNA polymerase moves along the DNA, unwinding it and adding new nucleotides, one at a time, to the 3′ end of the growing RNA strand; and

3. termination, the recognition of the end of the transcription unit and the separation of the RNA molecule from the DNA template.

Prokaryotic RNA Polymerase

There is only one type of RNA polymerase in prokaryotes, which catalyzes the synthesis of all classes of bacterial RNA: mRNA, tRNA, and rRNA. Bacterial RNA polymerase (subunit composition shown in Fig. 2) is a large, multimeric enzyme **of 480 kD Molecular Weight.**

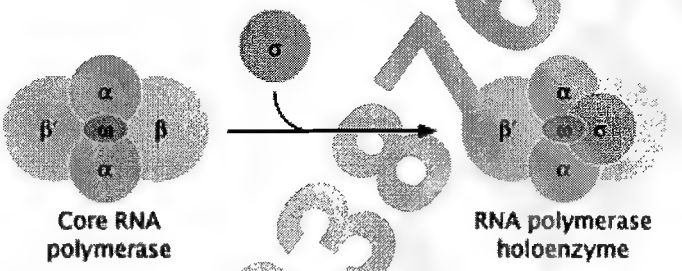


Figure 2: subunit composition of bacterial RNA polymerase

Bacterial RNA polymerases are composed of five subunits of polypeptide chains (Fig. 2) that make up the **core enzyme**: two copies of a subunit called alpha (α) and single copies of subunits beta (β), beta prime (β′), and omega (ω). The core enzyme catalyzes the elongation of the RNA molecule by the addition of RNA nucleotides.

The **sigma (σ) factor** is not a part of the core enzyme but it controls the binding of RNA polymerase to the promoter. It is a part of the **holoenzyme**. After sigma has associated with the core enzyme (forming a **holoenzyme**), RNA polymerase binds stably only to the promoter region and initiates transcription at the proper start site. Sigma is required only for promoter binding and initiation. When a few RNA nucleotides have been joined together, sigma usually detaches from the core enzyme.

Thus, the bacterial RNA Polymerase structure has:

1. α – Two sub units: Involved in making the pentameric core, comprising of α_2 β β′ω.

2. β – Contains the rNTP binding site

3. β′– Contains the template DNA binding site

4. ω – Stabilises the catalytic center for polymerization.

5. σ – Contains the promoter sequence-binding site. The role of this sub unit is in recognizing and binding to the promoter site. Once the nascent transcript reaches the stage of 8-9 nucleotides, σ sub-unit leaves the polymerase complex. However, this sub-unit is essential in ensuring that transcription starts at specific sites. If σ sub-unit is not present, transcription initiates randomly and cannot proceed beyond 20 nucleotides.

Many bacteria possess multiple types of sigma. *E. coli*, for example, possesses sigma 28 (σ^{28}), sigma 32 (σ^{32}), sigma 54 (σ^{54}), and sigma 70 (σ^{70}), named on the basis of their molecular weights. Each type of sigma initiates the binding of RNA polymerase to a particular set of promoters. For example, σ^{32} binds to promoters of genes that protect against environmental stress, σ^{54} binds to promoters of genes used during nitrogen starvation, etc.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Initiation and elongation of RNA chains

Initiation comprises all the steps necessary to begin RNA synthesis, including

1. promoter recognition,
2. formation of the transcription bubble,
3. creation of the first bonds between rNTPs, and
4. escape of the transcription apparatus from the promoter.

Transcription initiation requires that the transcription apparatus recognize and bind to the promoter.

Promoters in prokaryotes

More than 100 promoter sequences in prokaryotes have been characterized. They differ from each other in some respects, but certain fundamental structures are well conserved. They are (as shown in Fig. 3):

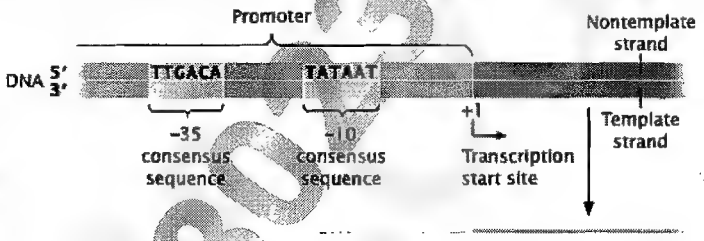


Figure 3: general design of prokaryotic gene promoters

1. **-10 sequence**: occurs at about 10 nucleotides before the point of transcription initiation. It contains a **consensus sequence** of TATAAT. It is also called **Pribnow box** (after its inventor R. Pribnow). The -10 sequence, being rich in AT is **the site of DNA duplex unwinding.**
2. **-35 sequence**: occurs at about 35 nucleotides before the point of transcription initiation. It contains a consensus sequence of TTGACA. It is also called the **recognition sequence**, because **this is where sigma factor binds**.

The distance between -10 and -35 sequences is highly conserved in *E.coli.* It is never less than 15 nt and more than 20 nt in length.

Some bacterial promoters contain a third consensus sequence - **upstream element** - that contains a number of A–T pairs and is found at about –40 to –60. The alpha subunit of the RNA polymerase interacts directly with this upstream element, greatly enhancing the rate of transcription in those bacterial promoters that possess it.

Initiation & Elongation

The process of RNA synthesis in bacteria involves first the binding of the RNA polymerase (RNAP) holoenzyme molecule to the template at the promoter site to form a **preinitiation complex, or PIC.** Binding is followed by a conformational change of the RNAP, and the first nucleotide then associates with the initiation site on the β subunit of the enzyme. In both prokaryotes and eukaryotes, a purine ribonucleotide is usually the first to be polymerized into the RNA molecule. In the presence of the appropriate nucleotide, RNAP catalyzes the formation of a phosphodiester bond, and the nascent chain is now attached to the polymerization site on the β subunit of RNAP. (See Figure 4.)

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

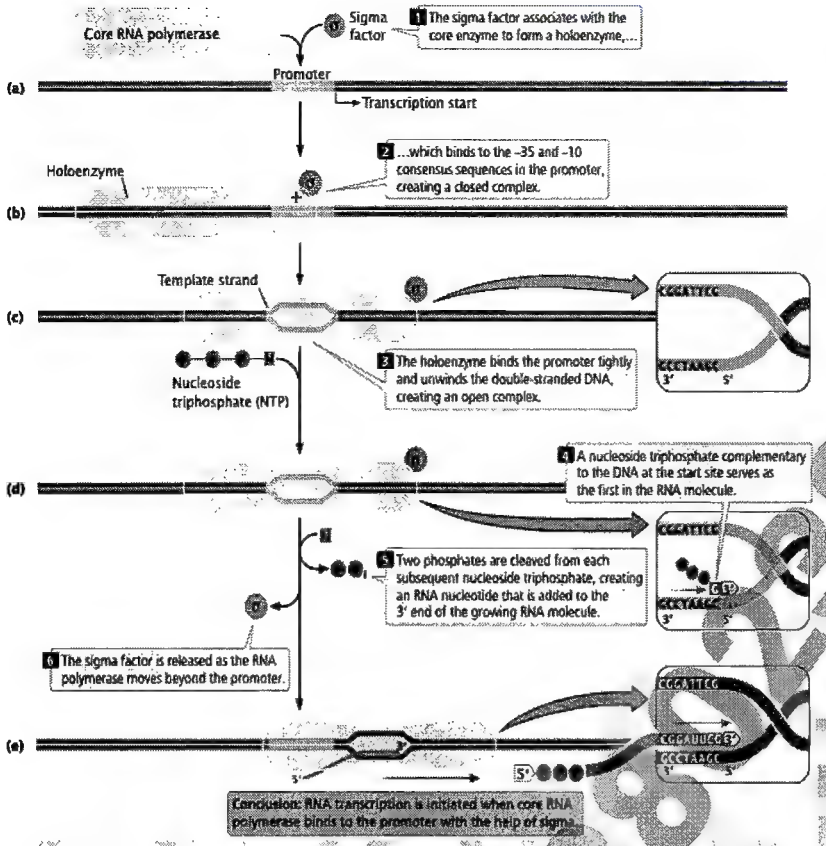


Figure 4: Initiation of prokaryotic trancription

Initiation of formation of the RNA molecule at its 5' end then follows, while elongation of the RNA molecule from the 5' to its 3' end continues antiparallel to its template. The enzyme polymerizes the ribonucleotides in a specific sequence dictated by the template strand and interpreted by Watson-Crick base-pairing rules. Pyrophosphate is released in the polymerization reaction. This pyrophosphate (PP_i) is rapidly degraded to 2 mol of inorganic phosphate (P_i) by ubiquitous pyrophosphatases, thereby providing energy for irreversibility on the overall synthetic reaction. As with eukaryotes, 5' triphosphate of this first nucleotide is maintained in prokaryotic mRNA.

Often in the course of initiation, RNA polymerase repeatedly generates and releases short transcripts, from 2 to 6 nucleotides in length, while still bound to the promoter. This stage is called Abortive Initiation Stage. After several abortive attempts, the polymerase synthesizes an RNA molecule from 10 to 20 nucleotides in length, which allows it to transition to the elongation stage.

After 10–20 nucleotides have been polymerized, RNAP undergoes a second conformational change leading to **promoter clearance.** A little before this stage i.e. after 8-9 nucleotides have been added to the transcript, the sigma subunit leaves the RNAP complex. Once this transition occurs, RNAP physically moves away from the promoter, transcribing down the transcription unit, leading to the next phase of the process, elongation.

As elongation proceeds, the polymerase holoenzyme unwinds and rewinds the DNA duplex. At any given moment the transcription bubble is not more than 20 nucleotides. During elongation, the RNAP core carries out the following steps cyclically.

1. DNA Unwinding
2. Nucleotide addition
3. Proof reading
4. Template-Transcript Separation

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

5. Reannealing of the separated DNA strands.

During elongation on an average, 40 nucleotides are synthesized per second. Elongation process is assisted by elongation factors GRE and Nus. These factors also help in proof reading.

Termination of transcription

Transcription terminates at well-defined **termination signals,** specific DNA sequences. With termination, the polymerase complex dissociates and the nascent transcript is released. There are two mechanisms of termination:

1. Rho [ρ] protein dependent termination
2. Rho independent termination.

1. In **ρ** dependent termination, a hexameric protein designated as Rho is employed. The precise mechanism is not well understood yet. According to Matsushuke [2001], the DNA template contains a 72 nt long signal sequence. When transcribed in to RNA molecule this sequence makes the binding and action site for the Rho protein. Earlier in 1996, Davidson had given the **Hot Pursuit Model** of Rho Dependent Termination according to which this protein moves behind the transcription apparatus. Rho is an ATP-dependent RNA-stimulated helicase that disrupts the nascent RNA-DNA complex at its binding and action site.

2. In Rho independent termination (mechanism shown in Fig. 5), there is a G-C rich termination signal involved in this mechanism. The overall signal sequence is such structured that once it is transcripted into RNA, it leads to hairpin type loop formation in the latter. The hairpin forms within 1^{-10} seconds and releases the nascent transcript from the template DNA molecule.

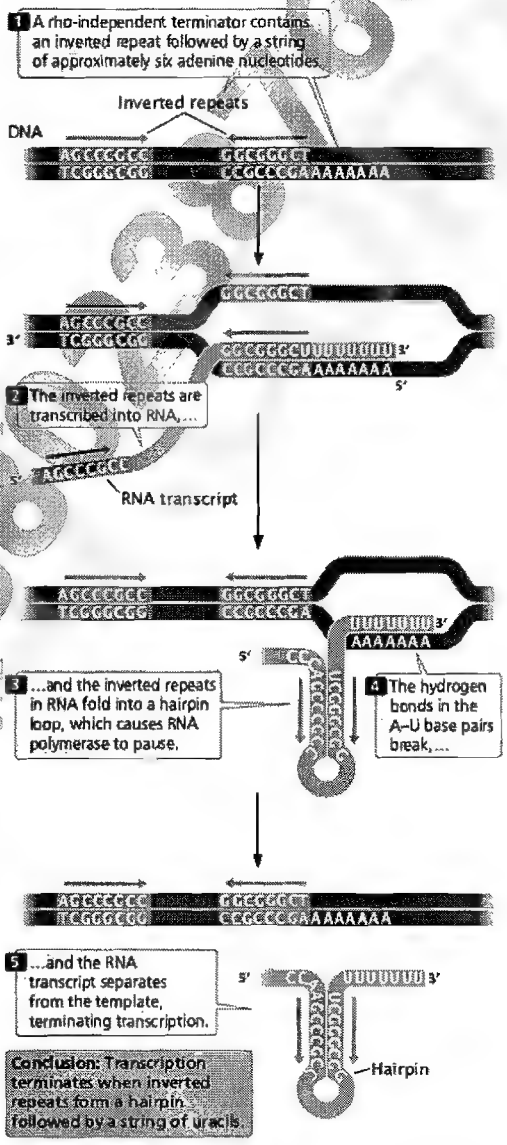


Figure 5: Rho independent termination

Transcription in eukaryotes

Transcription in Eukaryotic Cells: It's Distinctiveness from Transcription in the Prokaryotic Cells

Transcription in eukaryotic cells involves the same four stages described earlier for the prokaryotes, but the process in eukaryotes is more complicated than that in prokaryotes. The main differences are as follows:

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

1. ***Three different RNA polymerases*** transcribe the nuclear DNA of eukaryotes. Each synthesizes one or more classes of RNA.

2. ***Eukaryotic promoters*** are more varied than prokaryotic promoters. Furthermore, some eukaryotic promoters are actually located *downstream* from the transcription start point.

3. Binding of eukaryotic RNA polymerases to DNA requires the participation of additional proteins, called ***transcription factors***. Unlike the bacterial sigma factor, eukaryotic transcription factors are not part of the RNA polymerase molecule. Rather, some of them must bind to DNA *before* RNA polymerase can bind to the promoter and initiate transcription.

4. The eukaryotes show two functional classes of transcription factors:

    a. The class of factors that are essential for the transcription of all genes transcribed by an RNA polymerase - called the **basal transcription factors**;

    b. The **regulatory class of transcription factors**, which selectively act on specific genes.

5. ***Protein-protein interactions*** are very important in the first stage of eukaryotic transcription. Although some transcription factors bind directly to DNA, many attach to other proteins – either to other transcription factors or to RNA polymerase itself.

6. ***RNA cleavage*** is more important than the site where transcription is terminated in determining the location of the 3' end of the RNA product.

7. Newly forming eukaryotic RNA molecules typically undergo extensive ***RNA processing*** (chemical modification) both during and, to a larger extent, after transcription.

RNA Polymerases in Eukaryotes

The table below summarizes some of the properties of the three RNA polymerases that function in the nucleus of the eukaryotic cell, as well as two others found in mitochondria and chloroplasts. The nuclear enzymes are designated **RNA polymerases I, II, and III**. *In the recent years, new nuclear RNA polymerase identified - **RNA Polymerase IV** - found only in plants and responsible for synthesis of RNA involved in gene silencing.*

RNA Polymerase	Location	Main Products	α-Amanitin Sensitivity
I	Nucleolus	Precursor for 28S rRNA, 18S rRNA, and 5.8S rRNA	Resistant
II	Nucleoplasm	Pre-mRNA and most snRNA	Very sensitive
III	Nucleoplasm	Pre-tRNA, 5S rRNA, and other small RNAs	Moderately sensitive*
Mitochondrial	Mitochondrion	Mitochondrial RNA	Resistant
Chloroplast	Chloroplast	Chloroplast RNA	Resistant

*In mammals.

RNA polymerase I resides in the nucleolus and is responsible for synthesizing an RNA molecule that serves as a precursor for three of the four types of rRNA found in eukaryotic ribosomes (28S rRNA, 18S rRNA, and 5.8S rRNA). Its association with the nucleolus is because the nucleolus is the site of ribosomal RNA synthesis and ribosomal subunit assembly.

RNA polymerase II is found in the nucleoplasm and synthesizes precursors to mRNA. In addition, RNA polymerase II also synthesizes most of the snRNAs (small nuclear RNAs) involved

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

in posttranscriptional RNA processing. Thus, polymerase II is responsible for the synthesis of the greatest variety of RNA molecules. The enzyme is extremely sensitive to α-amanitin, which explains the toxicity of this compound to humans and other animals.

RNA polymerase III is also a nucleoplasmic enzyme, but it synthesizes a variety of small RNAs, including tRNA precursors and the smallest type of ribosomal RNA, 5S rRNA. Mammalian RNA polymerase III is sensitive to α-amanitin, but only at higher levels of the toxin than are required to inhibit RNA polymerase II.

Structurally, RNA polymerases I, II, and III are somewhat similar to each other as well as to prokaryotic core RNA polymerase. The three enzymes are all quite large, with multiple polypeptide subunits and molecular weights around 500 Kd. RNA polymerase II, for example, has more than ten subunits of at least eight different types. The three biggest subunits are evolutionarily related to the prokaryotic RNA polymerase subunits a, b and b'. Three of the smaller subunits lack that relationship but are also found in RNA polymerases II and III. The RNA polymerases of mitochondria and chloroplasts resemble their prokaryotic counterparts closely, due to the probable origins of these organelles as endosymbiotic bacteria.

The Three Classes of Promoters Found in Eukaryotic Nuclear Genes: One for Each Type of rna Polymerase

The promoters to which eukaryotic RNA polymerases bind are even more varied than prokaryotic promoters, but they can be grouped into three main categories, one for each type of polymerase. Figure below shows characteristic features of the three types of promoters.

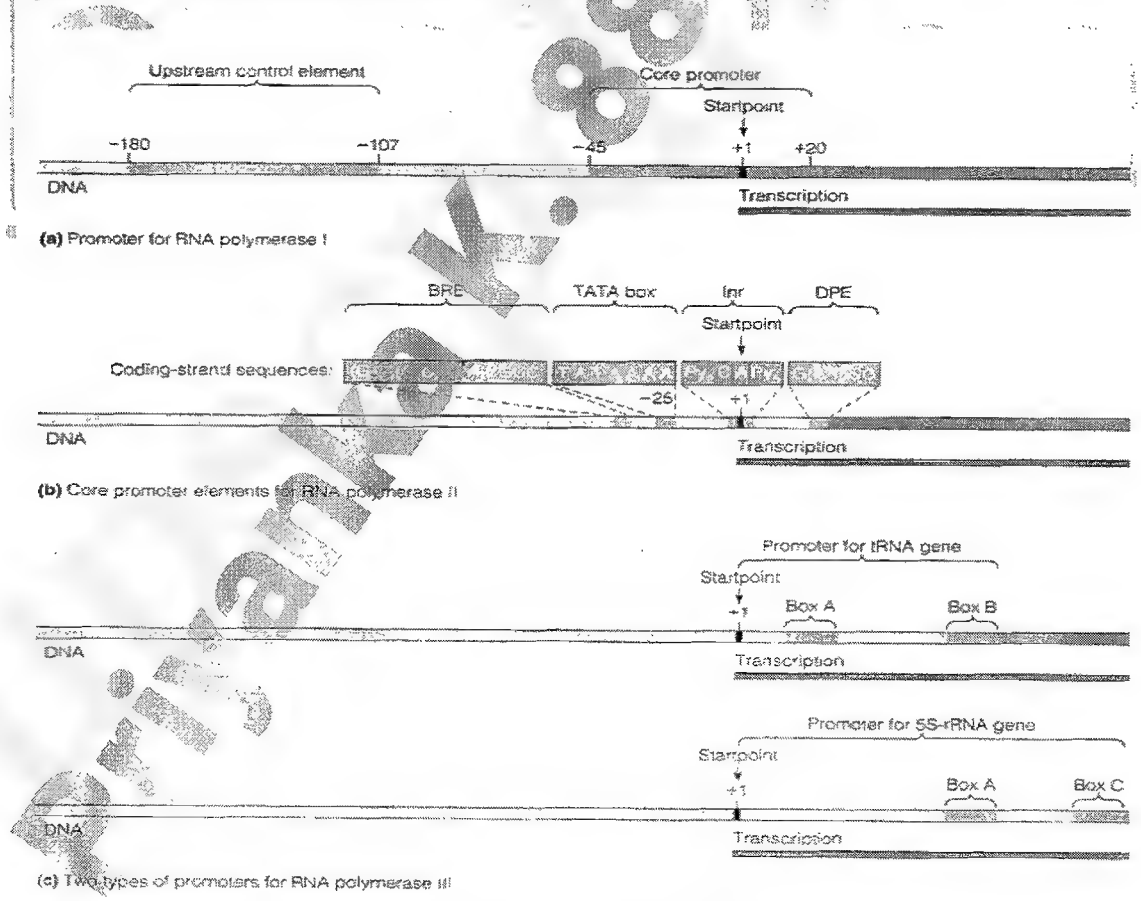


Figure 6: the various types of gene promoters in eukaryotes

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

1. **Promoter for RNA Polymerase I:** The promoter used by RNA polymerase I has two parts. The part called the **core promoter**-defined as the minimal set of DNA sequences sufficient to direct the accurate initiation of transcription by RNA polymerase – actually extends into the nucleotide sequence to be transcribed. It is sufficient for proper initiation of transcription, but transcription is made more efficient by the presence of an **upstream control element**. Attachment of transcription factors to both parts of the promoter facilitates the binding of RNA polymerase I to the core promoter and enables it to initiate transcription at the start point.

2. **Promoter for RNA Polymerase II:** In the case of RNA polymerase II, at least four types of DNA sequences are involved in core promoter function. These four elements are

   i. a short **initiator (Inr)** sequence surrounding the transcription start point.

   ii. the **TATA box**, which consists of a consensus sequence of TATA, usually located about 25 nucleotides upstream from the start point;

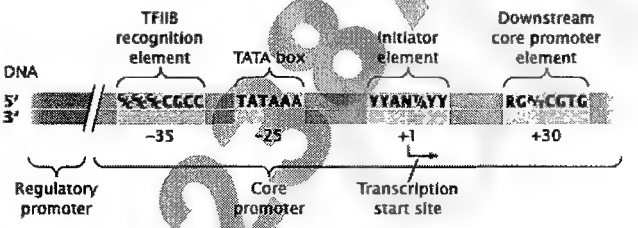


Figure 7: the four functional elements in the promoter for RNA polymerase type ii in eukaryotes

   iii. the **TFIIB recognition element (BRE)** located immediately upstream of the TATA box;

   iv. the **downstream promoter element (DPE)** located about 30 nucleotides downstream from the start point.

   These four elements are organized into two general types of core promoters:

   i. ***TATA-driven promoters***, which contain an *Inr* sequence and a TATA box with or without an associated BRE, and

   ii. ***DPE-driven promoters***, which contain DPE and *Inr* sequences, but no TATA box or BRE.

   *Most protein-coding genes have additional short sequences further upstream -**upstream control elements**- that improve the promoter's efficiency*. Some of these upstream elements are common to many different genes; examples include the ***CAAT box*** (consensus sequence GCCCAATCT in animals and yeasts) and the **GC *box*** (consensus sequence GGGCGG). The elements within 100-200 nucleotides of the start point are often called ***proximal control elements*** to distinguish them from ***enhancer*** elements, which tend to be farther away and can even be located downstream of the gene.

3. **Promoter for RNA Polymerase III:** In contrast to RNA polymerases I and II, the RNA polymerase III molecule uses promoters that are entirely *downstream* of the transcription unit's start point. The promoters used by tRNA and 5S rRNA genes are different, but in both cases, the consensus sequences fall into two blocks of about 10 bp each. The tRNA promoter has consensus sequences called *box A* and *box B*. The promoters for 5S rRNA genes have box A and *box* C. Not shown in the figure is a third type of RNA polymerase III promoter, an upstream promoter that is used for the synthesis of other kinds of small RNA molecules.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Initiation and elongation

Initiation in eukaryotes also involves three steps:

1. RNA polymerase holoenzyme binds to the promoter region
2. A localized unwinding of the DNA duplex occurs
3. Synthesis of nascent RNA begins

Two broad classes of DNA sequences are important for the initiation of transcription: **promoters** and **enhancers**. A promoter is always found adjacent to (or sometimes within) the gene that it regulates and has a fixed location with regard to the transcription start point.

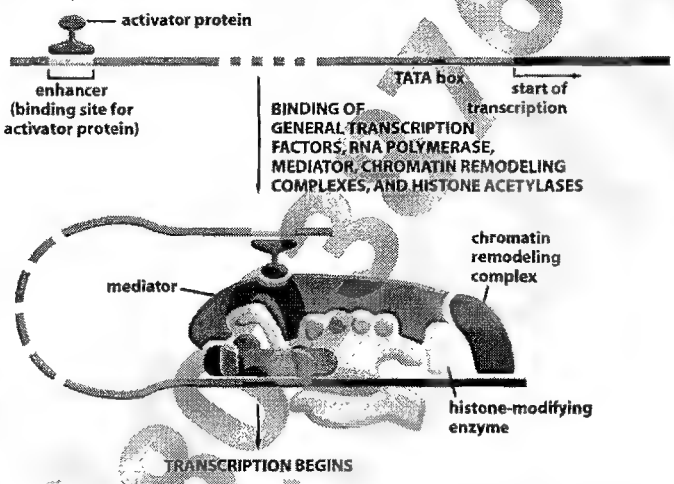


Figure 8: General picture of pre-transcription set up in eukaryotes

An enhancer, in contrast, need not be adjacent to the gene; enhancers can affect the transcription of genes that are thousands of nucleotides away, and their positions relative to start sites can vary. The enhancers bind to specific transcription activators and exert their influence by DNA bending and mediator proteins (Fig. 8).

In the initiation of eukaryotic transcription, two important classes of accessory proteins are required.

One class of accessory proteins - **general transcription factors** - along with RNA polymerase, form the **basal transcription apparatus** that assembles near the start site and is sufficient to initiate minimal levels of transcription. Another class - **transcriptional activator proteins** - bind to specific DNA sequences like enhancers and bring about higher levels of transcription by stimulating the assembly of the basal transcription apparatus at the start site.

The functioning of RNA polymerase II is dependent on a large number of protein molecules. They are collectively termed **transcription factors.** They are named in the series of **TF II** [Transcription Factor for RNA Pol **II**]. Initiation usually takes place in the following manner. Please see Figure 9.

1. In promoters containing a TATA box, **TF-IID** first binds to this promoter element. TFIID is a multiprotein complex that in which only one polypeptide, TATA-binding protein **(TBP)** binds to the TATA box. The complex also contains other polypeptides known as TBP-associated factors. In mammalian cells, TBP binds to the TATA box first and it is then joined by at least eight TBP associated factors to form TFIID. TBP is a monomeric protein. TBP has been shown to have a saddle structure by which it interacts with DNA in the minor groove. Thus, the inside of the saddle binds to DNA at the TATA box and the outside surface of the protein is available for interactions with other protein factors. Binding of TBP deforms the DNA producing a kink of about 45°.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

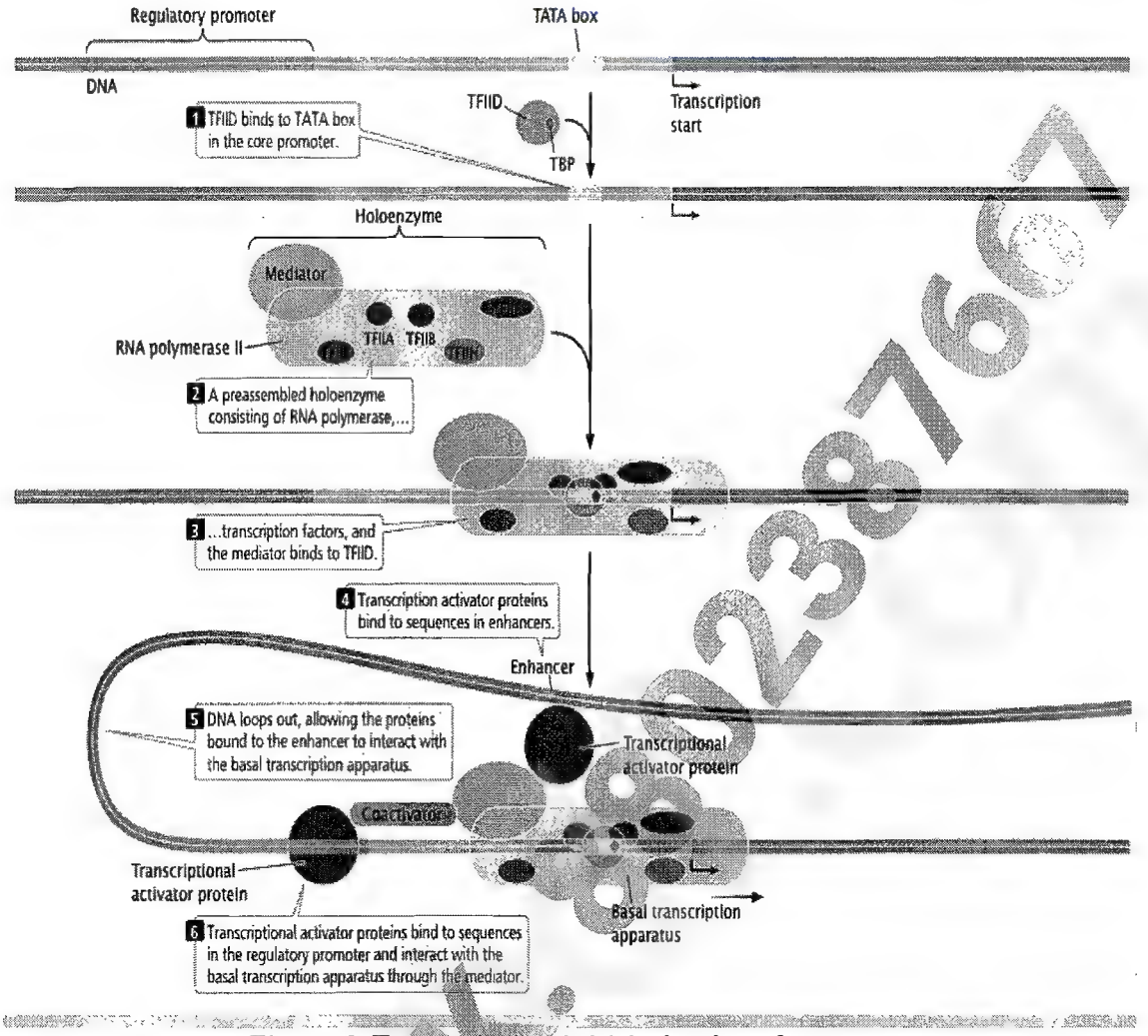


Figure 9: Transcription inititiation in eukaryotes

2. According to the currently accepted model (Bell & Gann, 2005), the RNA polymerase II, with a series of general transcription factors, and a complex of proteins known as the **mediator** binds to the promoter, where TFII D is already bound. The general transcription factors include TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH. Their functions include:

   - TFIIA binds to TFIID and enhances TFIID binding to the TATA box.
   - TFIIB acts as a bridging factor allowing recruitment of the polymerase to the complex together.
   - TFIIE activates the polymerase complex.
   - TFIIF has unwinding (Helicase) activity.
   - TFIIH enhances the processivity of the polymerase enzyme by phosphorylating the Carboxyl Terminal Domain [CTD] of the polymerase enzyme. **CTD** or **Carboxyl Terminal Domain** is the seven amino acid repeat that is found in multiple copies at the –C end of the largest sub-unit of RNA Pol-II. Its phosphorylation confers processivity to the Pol II enzyme complex.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Once the transcription begins, the RNA Pol II is sufficient to carry out elongation in 5'→3' direction. The elongation in eukaryotes occurs at a rate of ~20nt per second. Groups of scientists in March 2005 reported that elongation of transcription requires certain elongation factors like EF-IIS, hsPT5, Kinase TEFb and TAT-SF1.

Termination

There is no well-defined termination site in eukaryotes. However, there are **cleavage sites** where the nascent transcript is cut by endonuclease type enzymes. These enzymes are recruited by two proteins - **Cleavage and Polyadenylation Specificity Factor** (CPSF) and **Cleavage Stimulation Factor** (CStF). These proteins stay attached to the CTD of the RNA Polymerase II. But, at the time of cleavage these proteins get transferred onto the nascent transcript to recruit other proteins. The mechanism of nascent RNA cleavage is now well understood. The cleavage mechanism depends on a **cleavage signal sequence** in the nascent transcript that is 5'-AAUAAA-3'. Cleavage cuts the pre-mRNA into two pieces: the mRNA that will eventually encode the protein and another piece of RNA that has its 5' end trailing out of the RNA polymerase. After cleavage, all proteins and enzymes leave the pre-mRNA, except the CPSF, which now recruits the enzyme **Poly-A Polymerase** and initiates the **polyadenylation of the RNA chain**.

A unique feature of eukaryotic transcription termination is that the RNA polymerase II often continues to synthesize RNA hundreds or even thousands of nucleotides even after the cleavage process though an enzyme called **Rat1 nuclease** (Kim, Krogan *et al.*; 2008). Rat1 is a 5' →3' exonuclease - an enzyme capable of degrading RNA in a 5'→3' direction. Rat1 keeps digesting up the RNA as it moves. When Rat1 reaches the transcriptional machinery transcription terminates (Fig 10). Note that this mechanism is similar to that of rho-dependent termination in bacteria, except that rho does not degrade the RNA molecule.

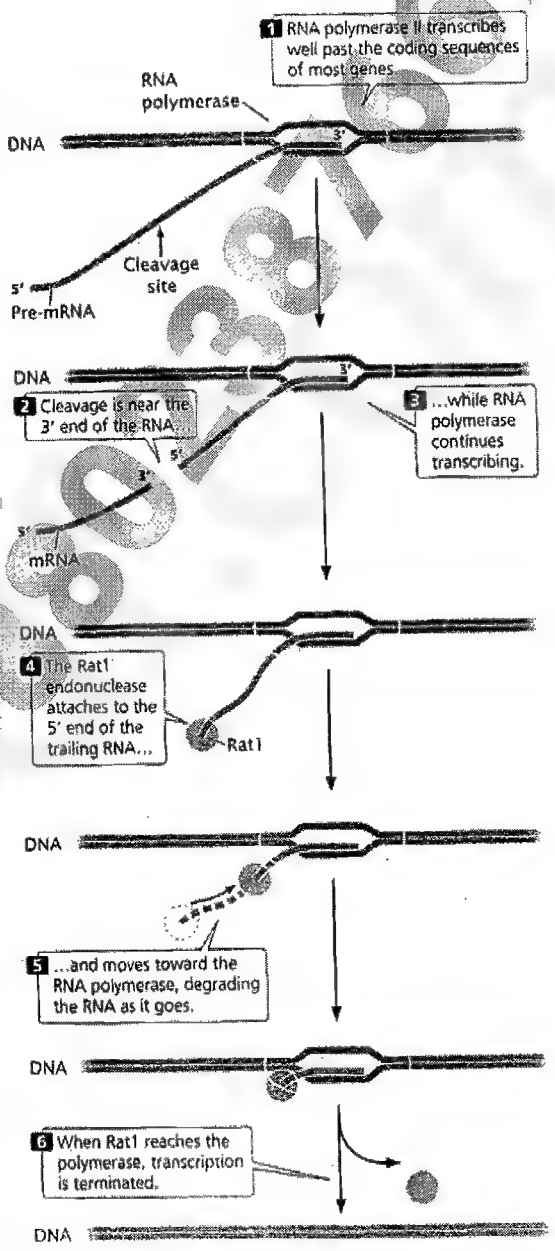


Figure 10: Transcription termination in eukaryotes for the genes transcribed by RNA polymerase II.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 3: Genetic Code

Introduction

During gene expression, DNA transfers information to mRNA in the form of a series of nucleotide triplet codes. Later in protein synthesis, ribosomes move along the mRNA molecule and polymerise the amino acids in accordance to the order of nucleotide triplet codes. These nucleotide triplet codes are called **genetic code**, because the genetic material (DNA) controls the synthesis of proteins via these codes only.

Discovery

The fist genetic code was identified by by Marshall Nirenberg and Heinrich Matthaei in 1961. They used a cell-free system to translate a poly-uracil RNA sequence (UUUUU...) and discovered that the polypeptide synthesized consisted of only the amino acid *phenylalanine*. They deduced that the codon UUU specified the amino acid phenylalanine. Subsequent work by Har Gobind Khorana identified the rest of the genetic code. In 1968, Khorana and Nirenberg received the Nobel Prize for their contribution.

Salient Features of the Genetic Code

1. **The code is a triplet codon.** The nucleotides of mRNA are arranged as a linear sequence of codons, each codon consisting of three successive nitrogenous bases. The concept of triplet codon has been supported by frameshift mutations.

2. **There are 64 codons.** Since RNA is constructed from four types of nucleotides, there are 64 possible triplet sequences (4^3). One of these codons (AUG) is the **start codon**, that is it specifies the point of peptide synthesis. In some cases, GUG can also act as the start codon. Three of these codons specify the termination of the polypeptide chain. They are called **stop codons** (UAA, UAG, UGA).

3. **The code is degenerate.** More than one codon may specify the same amino acid; this is called degeneracy of the code. For example, except for tryptophan and methionine, which have a single codon each, all other 18 amino acids have more than one codon. This arises because there are 61 codons to specify only 20 amino acids.

4. **The genetic code is almost universal.** The same codons are assigned to the same amino acids and to the same START and STOP signals in majority of genes in most organisms (Fig. 1) However, there are some exceptions in a few protists and organellar genes. These codes are called *non-standard codons*.

	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Figure 7: The Universal Genetic Code

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

5. **The code is non-overlapping**. In translating mRNA molecules the codons do not overlap but are read sequentially. Thus, a base in an mRNA is not used for different codons. However, it has been shown that in the bacteriophage $\phi \times 174$ there is a possibility of overlapping of genes and codons.

6. **The genetic code is commaless**, which means that no codon is reserved for punctuations. It means that after one amino acid is coded, the second amino acid will be automatically coded by the next three letters and that no letters are wasted as the punctuation marks.

7. **The code is non-ambiguous.** Non-ambiguous code means that a particular codon will always code for the same amino acid. In case of ambiguous code, the same codon could have different meanings at a time (codon) from the 5' end to the 3' end.

8. **The code has polarity.** The code is always read in a fixed direction, i.e., in the 5'→3' direction.

The non- standard codons

Organism	Codon	Should code for	Actually codes for
Mitochondrial genomes			
Mammals	UGA	Stop	Trp
	AGA, AGG	Arg	Stop
	AUA	Ile	Met
Drosophila	UGA	Stop	Trp
	AGA	Arg	Ser
	AUA	Ile	Met
Saccharomyces cerevisiae	UGA	Stop	Trp
	CUN	Leu	Thr
	AUA	Ile	Met
Nuclear and prokaryotic genomes			
Several protozoa	UAA, UAG	Stop	Gln
Candida cylindracea	CUG	Leu	Ser
Micrococcus sp.	AGA	Arg	Stop
	AUA	Ile	Stop
Euplotes sp.	UGA	Stop	Cys

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 4: An outline of pre-mRNA processing

In prokaryotic organisms, the primary transcripts of mRNA-encoding genes begin to serve as translation templates even before their transcription has been completed is because the site of transcription is not compartmentalized into a nucleus as it is in eukaryotic organisms. Thus, transcription and translation are coupled in prokaryotic cells. Consequently, prokaryotic mRNAs are subjected to no processing prior to carrying out protein synthesis.

On the other hand, nearly all eukaryotic RNA primary transcripts undergo extensive processing between the time they are synthesized and the time at which they serve their ultimate function, whether it be as mRNA or as a component of the translation machinery such as rRNA, 5S RNA, or tRNA or RNA processing machinery, snRNAs.

Processing occurs primarily within the nucleus and includes four major :

1. **Nucleolytic cleavage to smaller molecules** and coupled **nucleolytic and ligation reactions (splicing of introns).** The process is shown in Fig. 1.

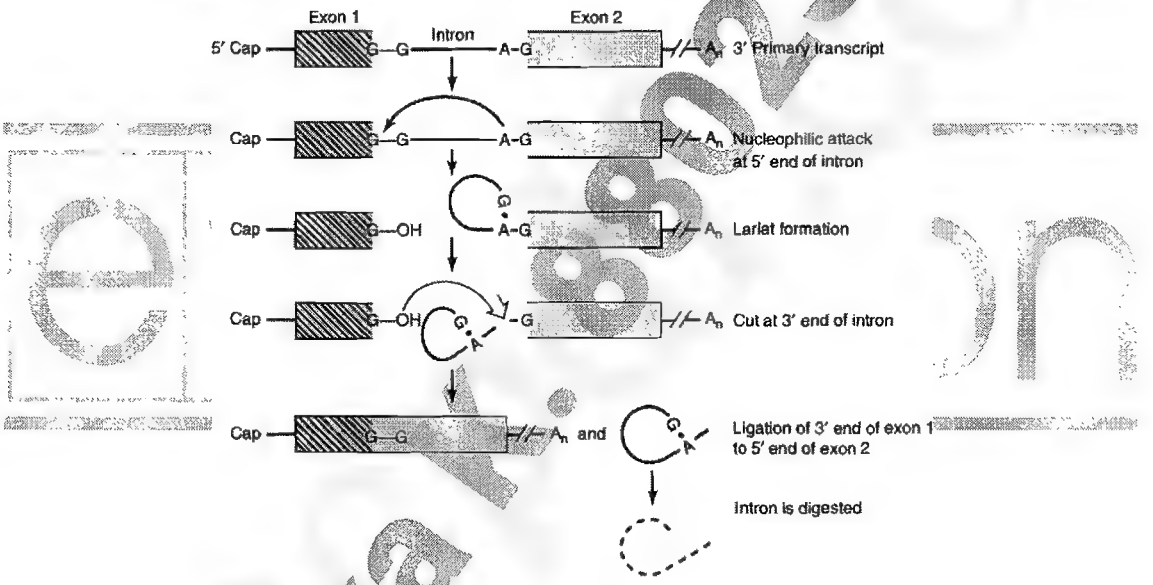


Figure 1: The processing of the primary transcript to mRNA. In this hypothetical transcript, the 5' (left) end of the intron is cut (↓) and a lariat forms between the G at the 5' end of the intron and an A near the 3' end, in the consensus sequence UACUAAC. This sequence is called the branch site, and it is the 3' most A that forms the 5'–2' bond with the G. The 3' (right) end of the intron is then cut (⇓). This releases the lariat, which is digested, and exon 1 is joined to exon 2 at G residues.

2. **Alternative Splicing Provides for Different mRNAs: The processing of hnRNA molecules is a site for regulation of gene expression.** Alternative patterns of RNA splicing result from tissue-specific adaptive and developmental control mechanisms. As shown above, the sequence of exon-intron splicing events generally follows a hierarchical order for a given gene. The use of alternative termination-cleavage-polyadenylation sites results in mRNA variability.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

3. **Messenger RNA Modification at the 5' & 3' Ends**: Eukaryotic mRNA molecules contain a 7-methylguanosine cap structure at their 5' terminal, and most have a poly (A) tail at the 3' terminal. The cap structure is added to the 5' end of the newly transcribed mRNA precursor in the nucleus prior to transport of the mRNA molecule to the cytoplasm. The **5' cap** of the RNA transcript is required both for efficient translation initiation and protection of the 5' end of mRNA from attack by 5' →3' exonucleases.

   Poly(A) tails are added to the 3' end of mRNA molecules in a posttranscriptional processing step. The mRNA is first cleaved about 20 nucleotides downstream from an AAUAAA recognition sequence. Another enzyme, poly(A) polymerase, adds a poly(A) tail. The **poly(A) tail** appears to protect the 3' end of mRNA from 3' →5' exonuclease attack.

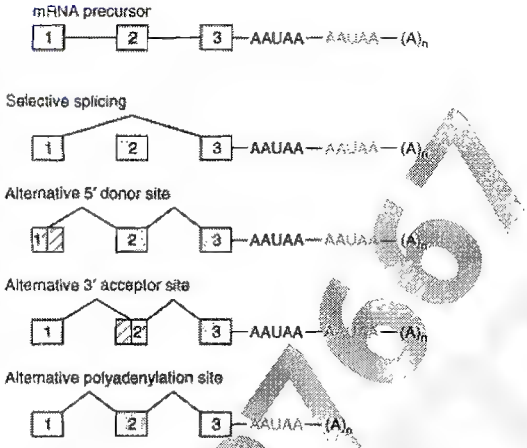


Figure 8: Mechanisms of alternative processing of mRNA precursors. This form of RNA processing involves the selective inclusion or exclusion of exons, the use of alternative 5' donor or 3' acceptor sites, and the use of different polyadenylation sites.

4. **RNA Editing** Changes mRNA after Transcription: The central dogma states that for a given gene and gene product there is a linear relationship between the coding sequence in DNA, the mRNA sequence, and the protein sequence. Changes in the DNA sequence should be reflected in a change in the mRNA sequence and, depending on codon usage, in protein sequence. However, exceptions to this dogma have been recently documented. Coding information can be changed at the mRNA level by **RNA editing**. In such cases, the coding sequence of the mRNA differs from that in the cognate DNA. An example is the apolipoprotein B *(apoB)* gene and mRNA (In liver, the single *apoB* gene is transcribed into an mRNA that directs the synthesis of a 100-kDa protein, apoB100. In the intestine, the same gene directs the synthesis of the primary transcript; however, a cytidine deaminase converts a CAA codon in the mRNA to UAA at a single specific site. Rather than encoding glutamine, this codon becomes a termination signal, and a 48-kDa protein (apoB48) is the result. ApoB100 and apoB48 have different functions in the two organs.

Priyanka 9687

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 5: Ribosomes

Introduction to Ribosomes

The ribosomes are macro-molecular complexes of RNA and proteins – involved in Peptide Synthesis – found abundantly in the cell cytoplasm of both prokaryotes and eukaryotes. On an average, a bacterial cell contains about 25,000 ribosomes, which make about 25% of the total dry cellular mass. A eukaryotic cell has between 1–10 million ribosomes in its cytoplasm.

Chemically ribosomes are:

~ 75% RNA, known as ribosomal RNA (rRNA)

~ 25% proteins.

The role of the ribosome is to hold the complex of mRNA and tRNAs together during the translation stage of peptide synthesis, catalyze the peptide bonds formation between the amino acids and ensure the accuracy of protein synthesis.

Occurrence and Types

1. In prokaryotes and Archaea: 70S type. Always free floating in the cytoplasm mostly aggregated in the nucleoid region.
2. In eukaryotes: 80S type with three possible locations.
    a. Free floating in the cytoplasm
    b. Attached to the RER membrane. In fact, it is the association of the ribosomes to the ER membrane surface because of which a particular region of the ER comes to be designated as RER.
    c. In eukaryotic nucleus: The 80S type of Ribosomes are also found in small quantities (~2%) in the eukaryotic nuclei – for the purpose of *mRNA Surveillance*. mRNA surveillance is a nuclear process for the detection and destruction of mRNAs that contain premature termination codons by a process called nonsense-mediated decay. (M.R. Culbertson, 2003).
3. In eukaryotic organelles: Various types of ribosomes have been reported from the genetically semiautonomous eukaryotic organelles viz. mitochondria (Mt) and the chloroplasts (Cp). Most of the plastidial ribosomes are of 70S type and they always occur in the chloroplast stroma. Mitochondrial ribosomes show great diversity. For example, among fungi, the Mt ribosome is 77S type, the Mt ribosome in mammals is of 55S type and in insects, the Mt ribosome is of 60S type. Mt ribosome is always found in the Mt matrix.

Ribosomal Architecture

The sub-units

Ribosomes consist of two subunits, large and small, that fit together. Prokaryotes have 70S ribosomes, each consisting of a (small) 30S and a (large) 50S subunit, whereas eukaryotes have 80S ribosomes, each consisting of a (small) 40S and a bound (large) 60S subunit.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

The size of ribosomes and the RNA molecules within them have traditionally been measured by their sedimentation coefficient in **Svedberg Units, denoted by S**. This is the rate at which the particles sediment in a gradient, under centrifugal force, and takes into account the size, the shape, and density of the particle.

The ribosome's RNA and protein composition is summarized in the following diagram.

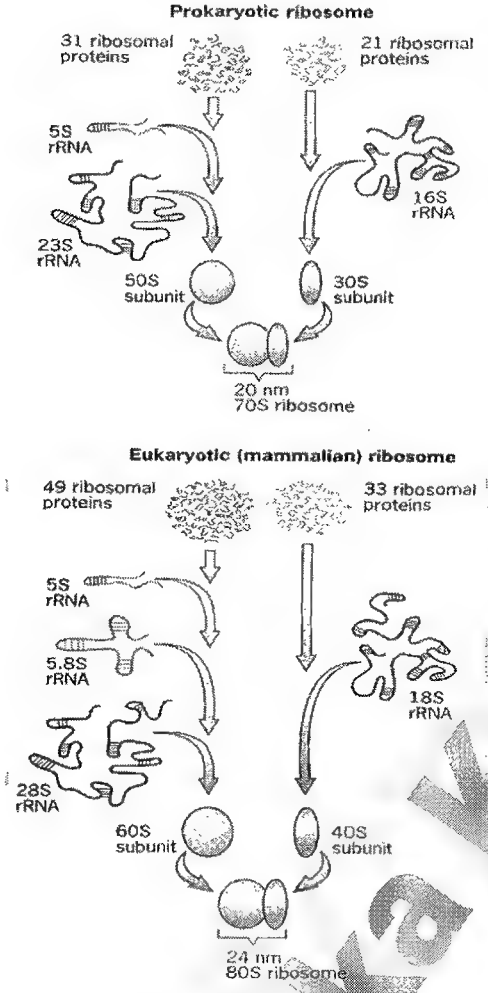


Figure 1: Ribosomal Subunits

Although ribosomes from prokaryotes and eukaryotes look similar in structure there are differences in the subunits and the composition of the proteins and rRNA within them. This makes protein synthesis an ideal target for antibiotics as drugs that affect bacterial ribosomes will not affect eukaryotic cells.

Molecular Anatomy and the Functional Domains of Ribosomes

Recently, four workers have separately unraveled the atomic structure of the ribosome. They are J. Frank (2000), N.Ban (2000), V.Ramakrishnan (2002) and P.B. Moore (2002). An EM structure was recently published by Mitra *et al.* (Nature, 2005).

V. Ramakrishnan is a scientist of Indian origin from the United States of America. He won the 2009 Nobel Prize for Chemistry for his pioneering work on the ribosomes. This breakthrough has led to a significant advancement in our current understanding about the structure and functions of the ribosomes.

Here we summarize our current knowledge about the ribosome's molecular anatomy and functional domains.

The rRNAs present in the ribosomes are crucial for both structure and function. Proteins only play a minor role in stabilizing the structure of the ribosomes.

1. The rRNAs are folded into well-defined conserved structures with many short duplex regions. Proteins interact with the RNA mostly at the surface level.

2. The major functional domains in the ribosome are as follows.

   a. ***The mRNA binding site*:** It is located in the smaller subunit's solitary rRNA, i.e. 16S in prokaryotes and 18S in eukaryotes.

   b. ***The A (aminoacyl) site*:** The larger part of this site is located in the larger subunit but a smaller part of it extends into the smaller subunit also. This is the site, where during peptide chain elongation a new charged tRNA arrives and binds the codon.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

c. ***The P (peptidyl) site:*** Like the A site, this site is also mainly located in the large subunit but a part of it extends into the smaller subunit also. This is where a charged tRNA is placed when the peptide bond has been formed between the newly arrived amino acid and the last amino acid of the pre-existing peptide chain. The tRNA is moved to this site by ribosomal translocation.

d. ***The E (exit) site:*** This is the site where a tRNA transiently binds the ribosome, before it leaves the complex after giving its amino acid. This site is almost entirely (~90%) located in the larger subunit of the ribosome.

e. ***The Peptidyl transferase centre:*** This is the catalytic domain of the ribosome. It is entirely located in the larger subunit. In the prokaryotes, this catalytic centre is a part of the 23S rRNA. In eukaryotes, the catalytic centre is made of 28S rRNA. Its 3-D structure brings the amino acids together and catalyzes the formation of a peptide bond between them. For the presence of this catalytic centre, ribosomes are regarded as a ribozyme.

f. ***The Peptide Exit tunnel:*** This is a tunnel composed of such rRNA domains, which cannot form any association with the nascent peptide chain. This tunnel occurs just above the P site and the newly growing peptide chain grows through this tunnel.

g. ***A factor binding centre:*** It occurs in the large sub-unit, near the A site. Some translation factors which bind to GTP interact with this domain. The factor binding centre helps these translation factors to hydrolyze their bound GTP into GDP.

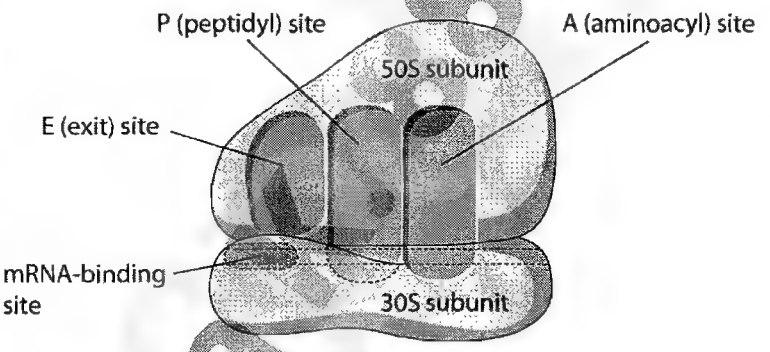


Figure 2: Some functional domains of ribosomes

Gross Morphology of Ribosomes

The currently accepted model of the ribosome morphology has been given by **James A. Lake** in 1981. Though initially given for the prokaryotic ribosome, this model is equally applicable to the eukaryotic ribosomes. It is summarized in Figure 3.

This completely asymmetrical model of ribosome suggested by James A. Lake (1981) essentially suggests the following.

1. The smaller subunit has
   a. A head
   b. A base
   c. A platform
2. The platform separates the head from the base by the help of a cleft. This cleft is an important functional region. It is suggested to be the site of codon-anticode interaction and as a part of binding site for initiation factors of protein synthesis.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

3. The large subunit consists of
    a. A ridge
    b. A central protuberance
    c. A stalk
4. The ridge and the central protuberance are separated with the help of a valley.

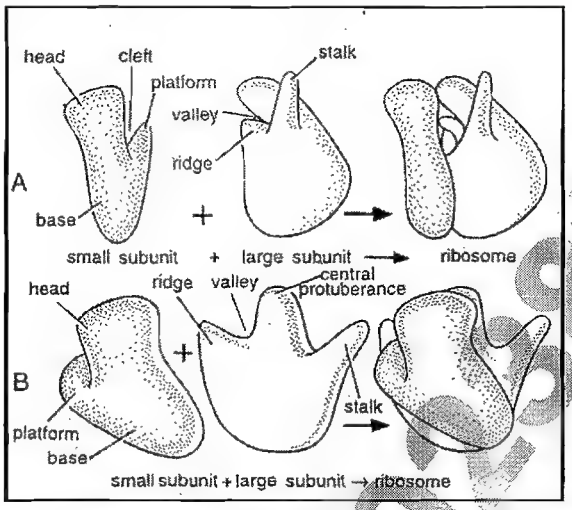


Figure 3: Gross Morphology of Ribosomes

Functions & Significance

1. **Role of ribosome in cellular metabolism:** With its precise structure and catalytic activity, a ribosome acts as an assembly line for protein synthesis in the cytoplasm. This is one of the final steps of the *central dogma*. The central dogma says that DNA is used to make RNA, which, in turn, is used to make protein. The DNA sequence in genes is copied into a messenger RNA (mRNA). Ribosomes then read the information in mRNA and use it to create proteins. This process is known as translation because ribosome "translates" the genetic information from RNA into proteins. Ribosomes do this by binding to an mRNA and using it as a template for the correct sequence of amino acids in a particular protein. The amino acids are attached to transfer RNA (tRNA) molecules, which enter one part of the ribosome and bind to the messenger RNA sequence. The attached amino acids are then joined together by another part of the ribosome. The ribosome moves along the mRNA, reading its sequence and producing a chain of amino acids (Fig 3).

    By playing the fundamental role in protein synthesis, ribosomes have a central role in cellular metabolism. Proteins are essential parts of all organisms and participate in virtually every process within cells. Therefore, they have been called *biomolecules of first importance* (Gr. *Proteus* = First).

    Many proteins are enzymes that catalyze biochemical reactions and are vital to metabolism. Proteins also have structural or mechanical functions, such as actin and myosin in muscle and the proteins in the cytoskeleton, which form a system of scaffolding that maintains cell shape. Other proteins are important in cell signaling, immune responses, cell adhesion, and the cell cycle.

    Thus, ribosomes provide for one of the most fundamental synthesis processes in a cell.

2. **Ribosomal RNA characteristics are important in medicine and in evolution**, as shown through the examples below.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

a. rRNA is the target of several clinically relevant antibiotics: Chloramphenicol, Erythromycin, Kasugamycin, Micrococcin, Paromomycin, Ricin, Sarcin, Spectinomycin, Streptomycin, and Thiostrepton.

b. 16S / 18S rRNA is the most conserved (least variable) RNA in all cells. For this reason, genes that encode the rRNA (rDNA) are sequenced to identify an organism's taxonomic group, calculate related groups, and estimate rates of species divergence.

3. **Ribosomes have a role in *mRNA Surveillance* and *Non-sense Mediated Decay Pathways*.** These pathways are nuclear processes for the detection and destruction of mRNAs that contain premature termination codons. They were discovered and described by M.R. Culbertson in 2003.

Priyanka K, 8802387661

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 6: Translation

Basic concepts

Translation refers to the complete process by which the genetic code carried by a particular molecule of mRNA is used to order and join Amino acids of a polypeptide chain.

Each amino acid is coded by between one and six codons and there are also codons that act as signals for the start and end of protein synthesis.

Exceptions for the universality of the genetic code exist in certain bacteria, unicellular eukaryotes such as *Acetabularia* and in organellar (such as mitochondrial) genomes.

Important features of the genetic code are:

1. The genetic code consists of a sequence of nucleotides in DNA or RNA. There are four letters in the code, corresponding to the four bases—A, G, C, and U (T in DNA).

2. The genetic code is a triplet code. Each amino acid is encoded by a sequence of three consecutive nucleotides, called a codon.

3. The genetic code is degenerate; that is, 64 codons encode only 20 amino acids in proteins. Some codons are synonymous, specifying the same amino acid.

4. Isoaccepting tRNAs are tRNAs with different anticodons that accept the same amino acid; wobble allows the anticodon on one type of tRNA to pair with more than one type of codon on mRNA.

5. The code is generally nonoverlapping; each nucleotide in an mRNA sequence belongs to a single reading frame.

6. The reading frame is set by an initiation codon, which is usually AUG.

7. When a reading frame has been set, codons are read as successive groups of three nucleotides.

8. Any one of three termination codons (UAA, UAG, and UGA) can signal the end of a protein; no amino acids are encoded by the termination codons.

9. The code is almost universal.

Conversion of the base sequence in the mRNA into an amino acid sequence involves RNA adapter molecules called transfer RNAs (tRNAs). These have a sequence of three bases, called the anticodon, which can base-pair with the codon sequence in the mRNA and, attached to 3' end of the tRNA molecule, is the amino acid corresponding to that codon. The process of matching up of tRNA molecules to the mRNA and the subsequent joining together of the amino acids into a polypeptide chain on the ribosomes - protein synthesis - is called translation (in cytoplasm).

The role of tRNAs

A tRNA molecule is a single strand of RNA (73-93 nucleotides long) folded into a cloverleaf secondary structure carrying the anticodon and amino acid binding site. It contains some unique

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

nucleotides, which are modified from the original A, U, G, and C nucleotides post-transcriptionally, including pseudouridine and inosine.

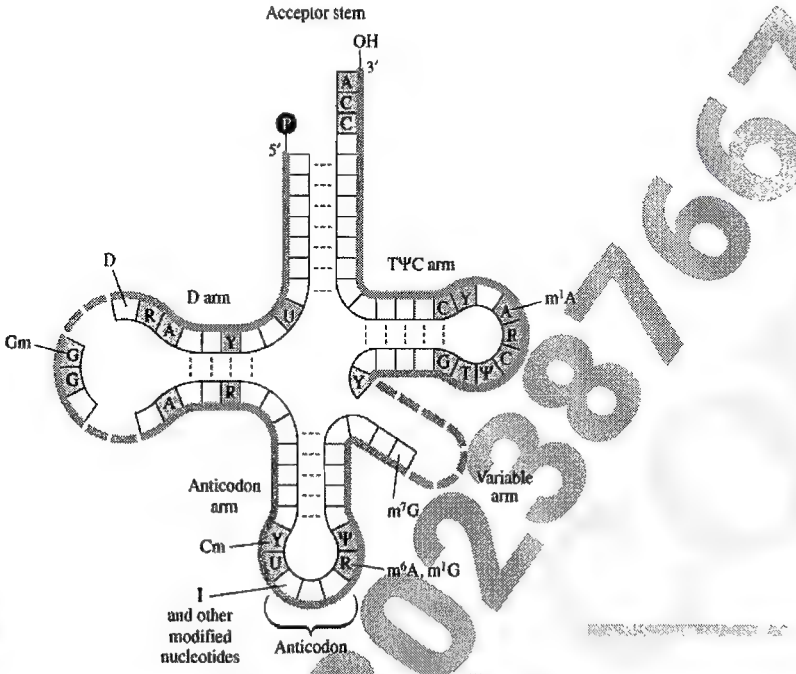


Figure 1: 2D structure of tRNA

There is at least one tRNA molecule for each amino acid in the cell but not necessarily one for each codon. The latter is made possible by wobble base-pairing i.e. non-Watson Crick base pairing at the 3rd base towards the 3' end of a codon. As a matter of fact, bacterial cells are known to have 40 – 50 different tRNAs which deal with the 61 codons.

The loading of the tRNA molecules with the correct amino acid is the primary requirement of protein synthesis. If the wrong amino acid is added to the tRNA molecule, it will subsequently be inserted incorrectly into the growing polypeptide chain, as the ribosome would not detect that the wrong amino acid was present. The enzymes responsible for the addition of the amino acid to the tRNA are aminoacyl-tRNA synthetases. There is at least one enzyme for each amino acid. The enzyme binds the correct amino acid and the tRNA molecule and catalyzes the addition of the amino acid to the 3' end of the tRNA molecule in two stages using ATP:

Amino acid + ATP → Aminoacyl-AMP + P~P

Aminoacyl-AMP + tRNA → Aminoacyl-tRNA + AMP

This process is sometimes called charging of the tRNA. The tRNA for an amino acid, called the cognate tRNA, is designated by a superscript (e.g. $tRNA^{Ala}$); once loaded with the amino acid it is written as Ala-$tRNA^{Ala}$.

The amino acid is bound to the 3'-terminal A residue with a high energy bond. This bond energy is later used to harness the process of peptide bond formation during the chain elongation.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Protein synthesis

Protein synthesis can be divided into three stages:

1. Initiation
2. Elongation, and
3. Termination.

At initiation the start codon for the protein is recognized: normally it is AUG which codes for methionine, but very occasionally it may be GUG (Valine). A complex is formed between the mRNA, the ribosome and the initiating tRNA.

During elongation, amino acids are added sequentially to the growing peptide chain in accordance with the codon sequence in the mRNA.

At termination, the end of the polypeptide chain is indicated by any of the three termination codons and the complex of mRNA, polypeptide, tRNA, and ribosomes breaks apart.

Each stage requires a number of different protein molecules, termed *factors* to ensure the correct order of events, which are different between prokaryotic and eukaryotic translation. The energy for protein synthesis is provided by the hydrolysis of GTP.

Prokaryotic Translation

Initiation

The first stage of protein synthesis is the binding of the small (30S) ribosomal subunit to the mRNA so that the first AUG codon is positioned in the P site. The correct positioning is achieved by base-pairing between a short purine rich sequence (consensus sequence: 5'-GGAGG-3'), called the *Shine-Dalgarno Sequence*, which is located in the mRNA, 8-13 nucleotides before the start site of translation, and a complementary sequence on the 16S RNA (consensus sequence 3'-CCUCC-5') on the small subunit of the ribosome. This mechanism, which is unique to prokaryotes, allows translation to start in the middle of a mRNA sequence as bacterial mRNAs frequently contain a number of genes which are translated independently.

The events of initiation proceed in the following manner.

1. Initiation factor 3 [IF3] binds to the small subunit

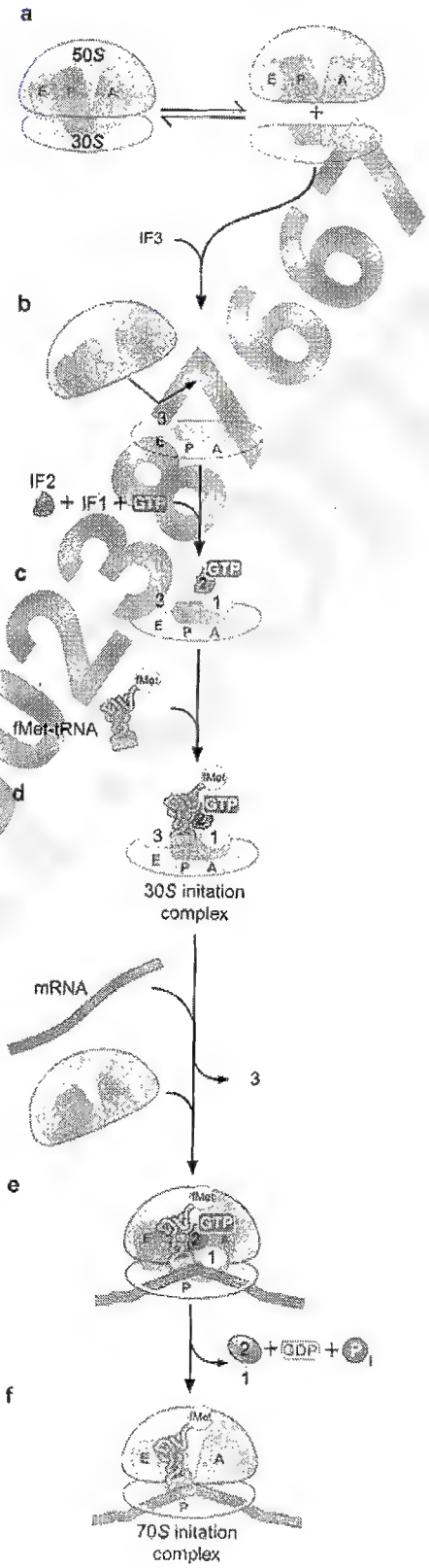


Figure 2: translation initiation in prokaryotes

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

of the ribosome and prevents it from binding to the large subunit of the ribosome. Due to this separation, the small ribosomal subunit can bind to mRNA and charged initiator tRNA (which carries *n-formyl methionine*). IF3 keeps the ribosomal subunits apart when protein synthesis is not going on.

2. Initiation factors 1 and 2 [IF1 and IF2] then bind to the small subunit of the ribosome. IF1 binds to the A site of the Ribosome and prevents the entry of the charged initiator tRNA at the A site. For successful translation initiation it is needed that the charged initiator tRNA enters the P site.

3. IF2 binds with GTP. Its role is in guiding the correct localization of the charged initiator tRNA.

Since the small sub-unit binds to the mRNA via the Shine Dalgarno sequence places the AUG initiation codon at the P site, hence the first tRNA, the initiator tRNA, can enter this site and its anticodon can base-pair with the AUG. The initiator tRNA is distinct from the tRNA which inserts methionine in response to AUG in internal sites in the polypeptide in two respects.

1. It carries N-formylmethionine (fMet) instead of methionine (Met);
2. It can directly enter the P site of the ribosome when the large subunit is absent.

Once the fMet-tRNA^{fMet} has bound to the AUG, the large (50S) ribosomal subunit joins to form the complete 70S ribosome. The small subunit of the ribosome aligned with the mRNA and charged initiator tRNA is called the 30S pre initiation complex.

With the joining of the larger subunit of the ribosome, first of all IF3 is released. Rest of the two initiation factors, IF1 and IF2 are released subsequently. When the large subunit binds, it activates a domain called *Factor Binding Centre*. This ribosomal domain interacts with all those translation factors which bind to GTP. The activation of the Factor Binding Centre stimulates the GTPase activity of IF2. Thus, GTP is hydrolysed into GDP. GTP hydrolysis changes the conformation of IF2, allowing it to leave the complex.

Just after IF2 release, the IF1 also leaves the complex.

Thus with the complete assembly of the ribosome, the 70S initiation complex is ready and Translation initiation is said to have accomplished.

Figure 3: Elongation and Translocation

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Elongation

Arrival of an appropriate charged tRNA at the A site begins elongation. Elongation is a cyclic process where one cycle is repeated every time when a new amino acid is added to the growing peptide chain. The elongation cycle has three components.

1. Charged mRNA binding at A site
2. Peptide bond formation between the newly arrived amino acid and the last added amino acid in the peptide chain
3. Translocation of the ribosome

The new arriving charged tRNA comes in a ternary complex of charged tRNA + EF-Tu + GTP. It binds to the A site by standard or Wobble codon-anticodon interaction. After the tRNA binds the A site, GDP and EF-Tu are released. The process of correct tRNA and mRNA base pairing is called tRNA accommodation. It is further enhanced by weak hydrogen bonds between rRNA and tRNA.

The activated form of EF-Tu and GTP are regenerated for the next elongation cycle by a process known as EF-Tu/EF-Ts Exchange Cycle.

The peptide bond formation is catalyzed by an inbuilt catalytic centre within the large ribosomal subunit - known as **Peptidyl Transferase** - a part of 23S rRNA in the ribosome's large sub-unit. So, this is actually a Ribozyme. This enzyme catalyzes the peptide bond formation between the C end of the AA bound to the tRNA at P site and the N end of the newly arrived AA at the A site. In effect this enzyme transfers the peptide chain formed so far to the newly entered amino acid.

Translocation (Fig 12) is fundamentally a process pushing the tRNAs already present within the ribosome by assistance of a protein factor designated as EF-G and energy from the hydrolysis of GTP. EF-G works by entering the A site of the ribosome.

During translation, the nascent polypeptide moves through a large, water-filled tunnel (approximately 10 nm x 1.5 nm) in the large subunit of the ribosome. The walls of this tunnel are made primarily of 23S rRNA. This structure provides a unique "low affinity" coating through which a polypeptide chain can easily slide.

The elongation cycle is operated repeatedly until the complex encounters a termination codon.

Termination

The end of the protein-coding message is signaled by the presence of one of three stop codons (UAA, UAG, or UGA). These are not recognized by a tRNA and do not specify an amino acid. Proteins known as **release factors** bind to any ribosome with a stop codon positioned in the A site, *forcing the peptidyl transferase in the ribosome to catalyze the addition of a water molecule instead of an amino acid to the peptidyl-tRNA.* This reaction releases the carboxyl end of the growing polypeptide chain from its attachment to a tRNA molecule, and since only this attachment holds the growing polypeptide to the ribosome, the completed protein chain is immediately released into the cytoplasm.

There are three types of release factors in prokaryotes falling into two categories.

a. **Class I release factors**, including RF 1 and RF 2, which actually release the newly synthesized peptide chain from the ribosome

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

b. **Class II release factor**, that is RF 3, a GTP hydrolyzing protein that enables the release of a class I release factor once it has performed its task in the ribosome.

First, any one of the two release factors, RF1 or RF2 binds the A site and then the peptide chain formed so far is released from the ribosome. Later with assistance of another release factor RF3, the class I release factor comes out of the ribosome.

Release factors are an example of **molecular mimicry**, whereby one type of macromolecule resembles the shape of a chemically unrelated molecule. In this case, the three-dimensional structure of release factors (made entirely of protein) resembles the shape and charge distribution of a tRNA molecule. This shape and charge mimicry helps them enter the A-site on the ribosome and cause translation termination.

Soon afterwards, the ribosome also disintegrates into its sub-units with help of Ribosome Recycling Factor, GTP and Elongation Factor G. The ribosome then releases the mRNA and separates into the large and small subunits. The disintegrated state of ribosome is then stabilized by IF3.

Eukaryotic Translation

In a eukaryotic cell two types of translation proceed.

1. Cytoplasmic translation, which is very different from the prokaryotic translation, especially in the process of initiation.

2. Organellar translation that occurs in the mitochondrial matrix and the chloroplast stroma and resembles the prokaryotic translation.

In this section, we shall focus on the cytoplasmic translation. Our understanding of the eukaryotic translation has undergone a major change after some recent works of V. RamaKrishnan [2002] and Zavialov, Ehrenberg *et al* [2005].

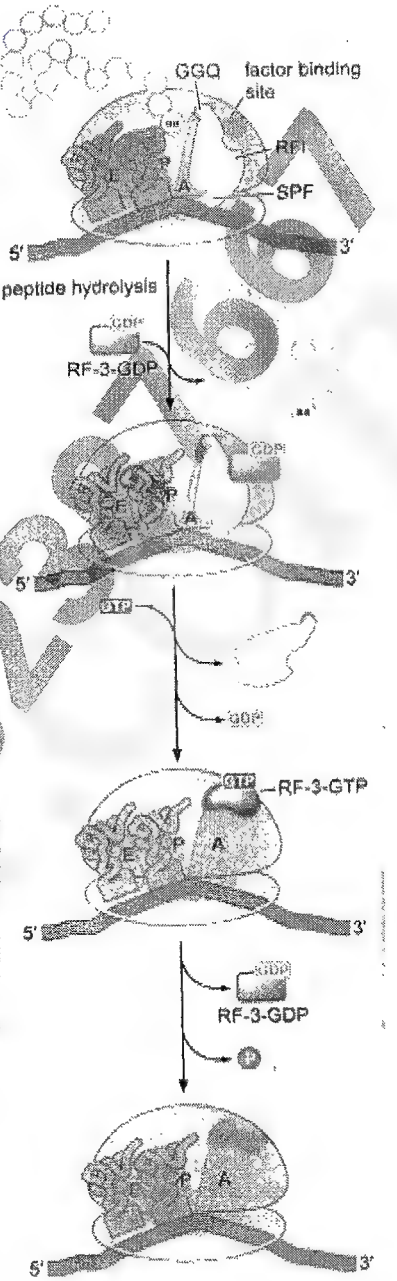


Figure 4: Termination

Translation in eukaryotes differs from the prokaryotes principally in the following respects:

1. The initiation is very elaborate and there are a greater number of initiation factors. The eukaryotic initiation factors are designated by the symbol of eIF.
2. No Shine Delgarno sequence to indicate the precise location of the initiation codon. Rather, there is a scanning mechanism employed by the protein synthesizing machinery.
3. The initiation codon is always AUG and the first Methionine coming in the peptide chain is not formylated.
4. A lesser number of release factors are involved; in most cases, there is just one or two release factors.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Initiation

There is no Shine Delgarno sequence. However, the identification of the initiation codon is made easy by its specific nucleotide surrounding. This is often referred to as the sequence context or the Kozak's sequence, after the name of its discoverer, Marilyn Kozak. The consensus part of the Kozak's sequence is 5'-ACCAUGG-3'.

Marilyn Kozak also gave a scanning model for the mechanism of locating the initiating codon, supported by experimental evidences. According to this model [after modifications by *RamaKrishnan, 2002 and Zavialov 2005*] translation initiation takes place the following way.

The smaller subunit of the ribosome [40S] is first bound by the eIF3, 1A. eIF3 prevents the binding by larger subunit, eIF1 blocks the A site.

The first triple binding facilitates the binding of the *ternary complex* of eIF2, the initiator tRNA and GTP. Along with this eIF5b also binds to the complex. eIF5b is the eukaryotic equivalent to IF2. This is because eIF5b interacts with eIF1 and also with the factor binding centre of the ribosome. In the correct placement of

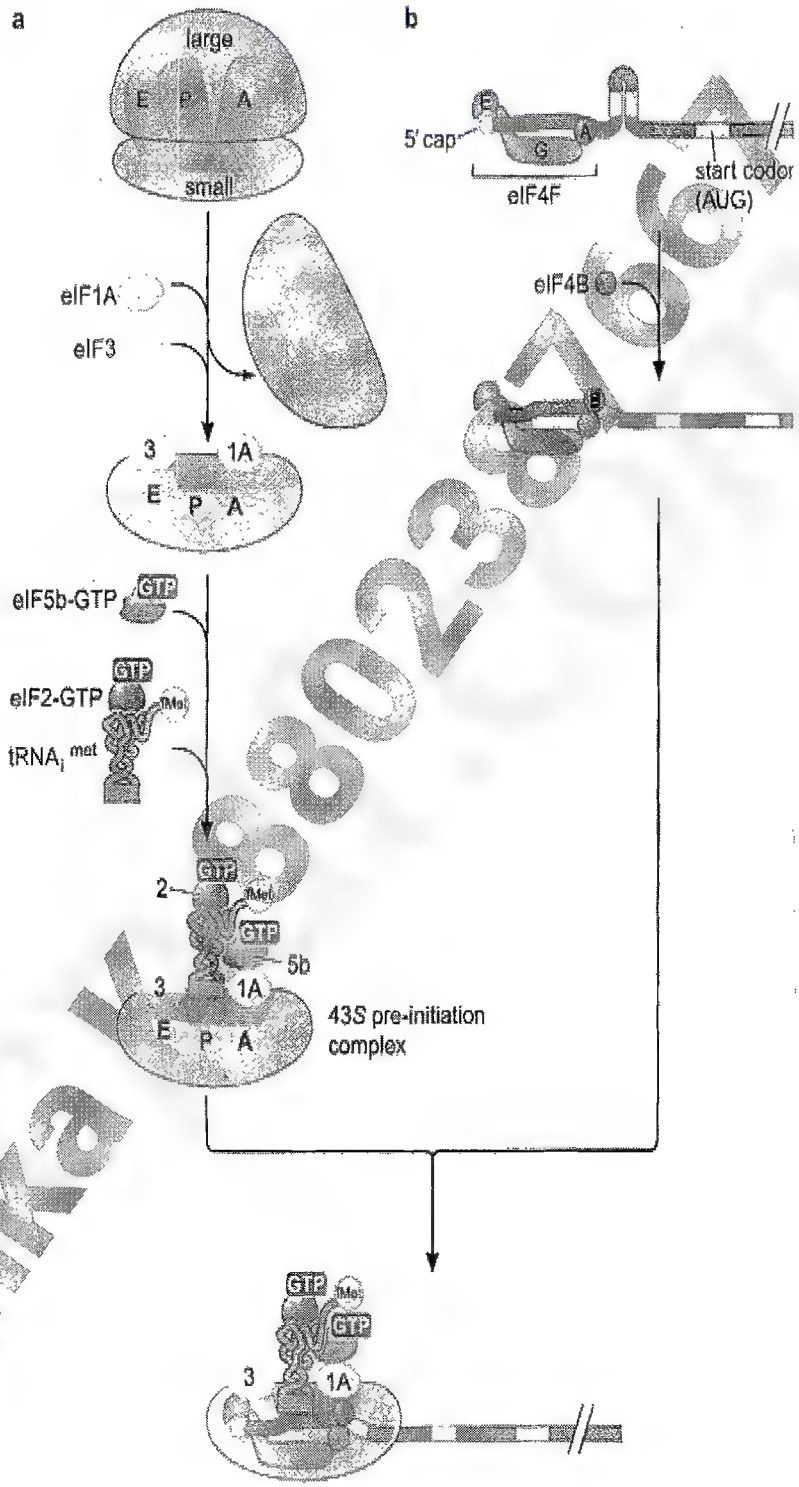


Figure 5: early stages of translation initiation in eukaryotes

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

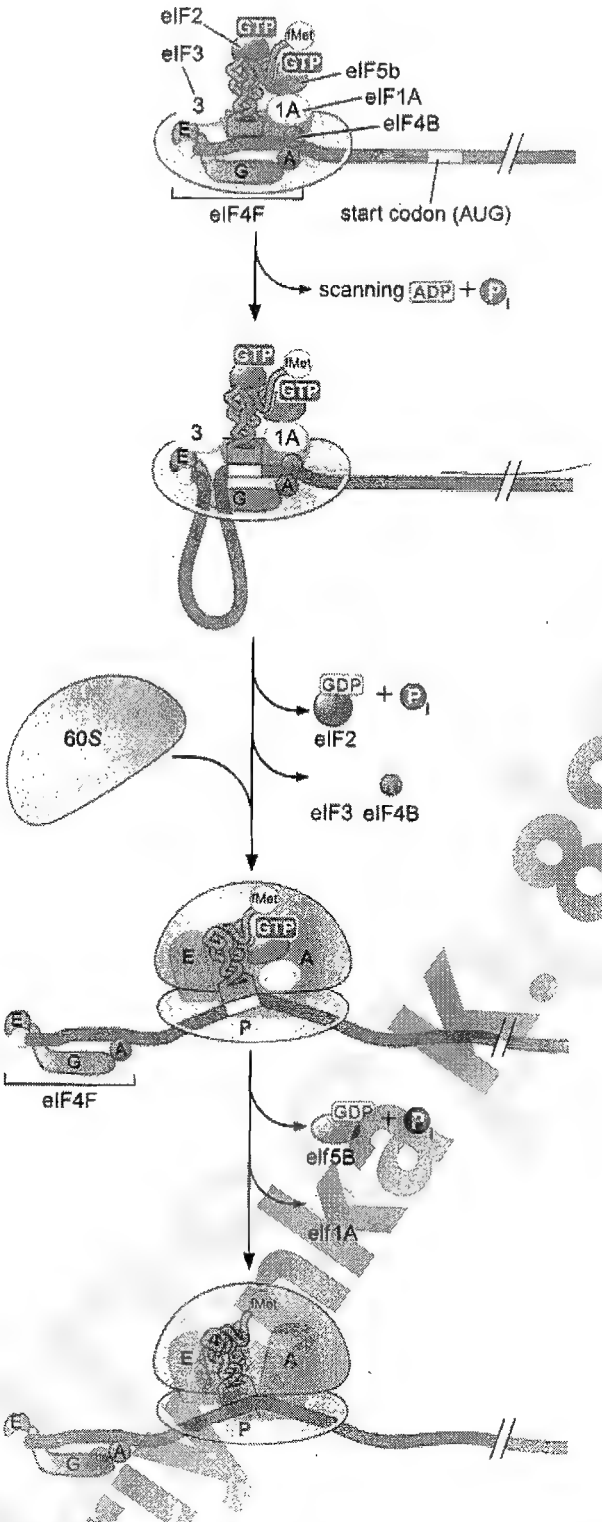


Figure 6: Scanning of mRNA during eukaryotic translation initiation

Centre. eIF1 is the last initiation factor to be released.

charged initiator tRNA both eIF2 and eIF5b play roles. After all this bindings, the 43S pre-initiation complex is ready.

This pre-initiation complex is ready to scan the mRNA for the Kozak sequence. The scanning process is shown below. Description follows later.

To resolve the secondary structure present within the mRNA and also to ease out any obstruction caused by the proteins bound to it, some initiation factors bind to the mRNA. These initiation factors belong to the eIF4F Complex with three sub-units: 4E, 4G and 4A. 4E is now known to bind to the 5'cap. 4E also facilitates the binding of 4G and 4A.

After the assembly of eIF4F Complex, the eIF4B binds next to the 4A subunit of the 4F complex.

The 4B factor has ATP dependent helicase activity that eases out the secondary structure present within the mRNA.

After this, the scanning process starts. ATP molecules are hydrolyzed to run it. eIF2B helps in the movement of the $tRNA_i$ along the mRNA molecule. Once the initiation codon is found within the Kozak's sequence, a Watson –Crick base pairing occurs between the AUG and the anticodon of the $tRNA_i$. This base pairing acts as a signal for the binding of the larger 60S ribosomal subunit.

The 60S subunit now binds with help of eIF6. A molecule of GTP is hydrolyzed to allow this binding and also to release all other initiation factors. eIF5b hydrolyses its GTP due to interaction with the Factor Binding

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

After the binding of the 60S subunit, the $tRNA_i$ gets precisely positioned at the P site. With this, the translation initiation is accomplished.

Around the same time mRNA makes a circle due to association between eIF4E and PolyA Binding Proteins (PABP) at 3' end.

Elongation

The process of elongation is very similar in prokaryotes and the eukaryotes. The role played by the prokaryotic EF-Tu is taken over by eEF1α. The roles of EF-Ts and EF-G are performed by eEF1-βγ and eEF-G respectively.

After the atomic structure elucidation of the ribosomes by Moore [2002], we now know that the peptidyl transferase centre is located within the 28S rRNA of the larger subunit. The growing peptide chain emerges out of the ribosome through a tunnel just above the P site.

Termination

Mostly, there are two release factors in the eukaryotes, designated as eRF1 (class I) and eRF3 (class II). The steps in translation termination are also similar in the prokaryotes and the eukaryotes.

After termination, similar to the prokaryotes, the ribosome also disintegrates into its 40S and 60S sub-units.

In both prokaryotes and eukaryotes, Ribosome Recycling Factor is essential to clean up ribosomes and make them available for the next round of Translation.

Priyanka K 8828161

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 7: Regulation of prokaryotic gene expression

Gene expression is the process by which a gene's information is converted into the structures and functions of a cell. Gene expression is a multi-step process (Fig. 1) that begins with transcription, post transcriptional modification and translation, followed by folding, post-translational modification and targeting.

Gene expression, its timing, location and extent, must tightly be controlled.

Regulation of gene expression (gene regulation) *is the cellular control of the amount and timing of appearance of the functional product of a gene. Although a functional gene product may be an RNA or a protein, the majority of the known mechanisms regulate the expression of protein coding genes.*

Any step of gene expression shown in Figure 1 may be controlled.

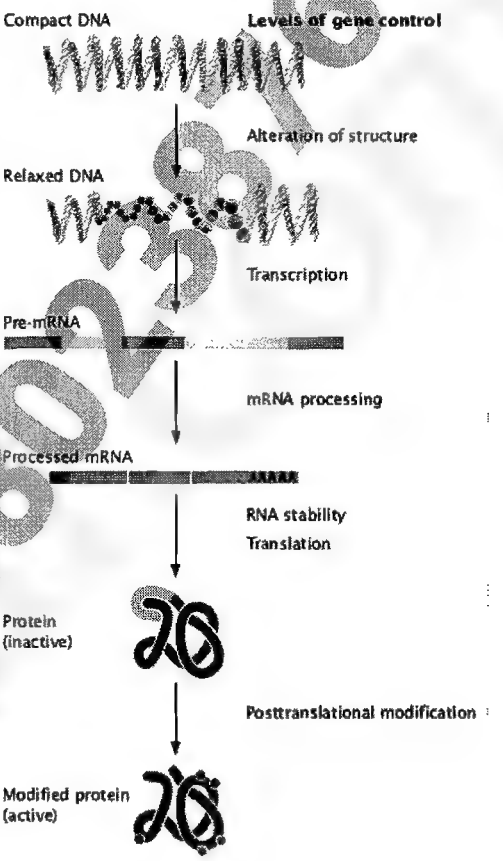


Figure 1: Steps in Gene Expression

Gene regulation has the following advantages.

1. The cell gets a control over structure and function by expressing only those genes which are required to be expressed at a particular location at a particular point of time.
2. The eukaryotes get the basis for cellular differentiation, morphogenesis and the organism has a control over various developmental events at different stages of the life cycle.
3. The versatility and adaptability (by way of biochemical flexibility) to any organism under varying environmental conditions is ensured by differential gene expression.
4. A great economy of ATP (or GTP), the cellular currency of free energy, is also ensured which would have otherwise been wasted in unnecessarily expressing all the genes all the times.
5. A possible prevention of metabolic accidents & metabolic chaos is also achieved by gene regulation (that would have otherwise resulted from expressing all the genes all the times).

Generally, there are two aspects of Gene Regulation.

1. **Spatial Gene Regulation** in which different sets of genes express in different parts of the body. This type of gene regulation applies to multicellular eukaryotes only where it plays a vital role in differentiation and morphogenesis.

2. **Temporal Gene Regulation** in which different sets of genes express at different points of time. This type of gene regulation applies to both prokaryotes and eukaryotes. It plays a vital

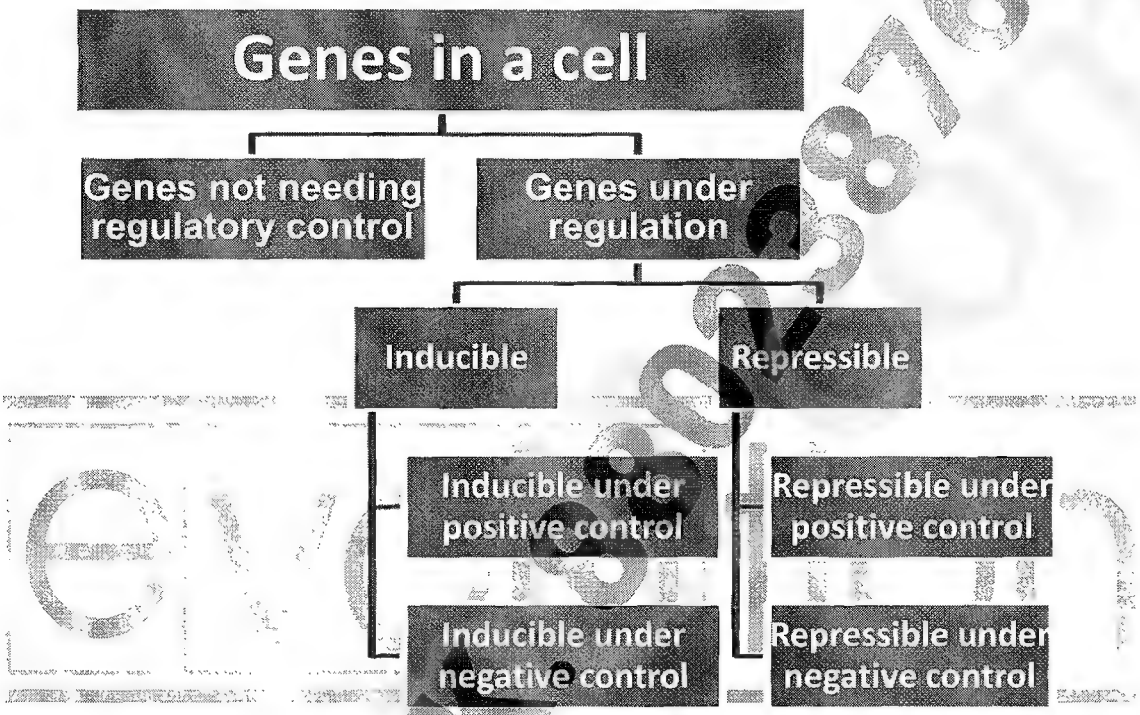
Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

role in the organism's adjustment strategies under various environmental conditions. It also mediates differentiation related to developmental stages.

Thus in bacteria, gene regulation maintains internal flexibility, turning genes on and off in response to environmental changes. In multicellular eukaryotic organisms, gene regulation brings about cellular differentiation.

Types of Genes from Regulatory View-point



A brief outline of the four major modes of gene regulation

As shown in the chart on the following page, there are four major types of gene control.

1. Positive control of inducible genes
2. Negative control of inducible genes
3. Positive control of repressible genes
4. Negative control of repressible genes

The regulation of gene expression-induction and repression can be accomplished by both positive control mechanisms and negative control mechanisms. Both mechanisms involve the participation of **regulator** genes. In **positive control mechanisms,** the product of the regulator gene is an activator whereas in **negative control mechanisms,** the product of the regulator gene is a repressor. Positive and negative regulations are illustrated for both inducible and repressible systems in Figure 2.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

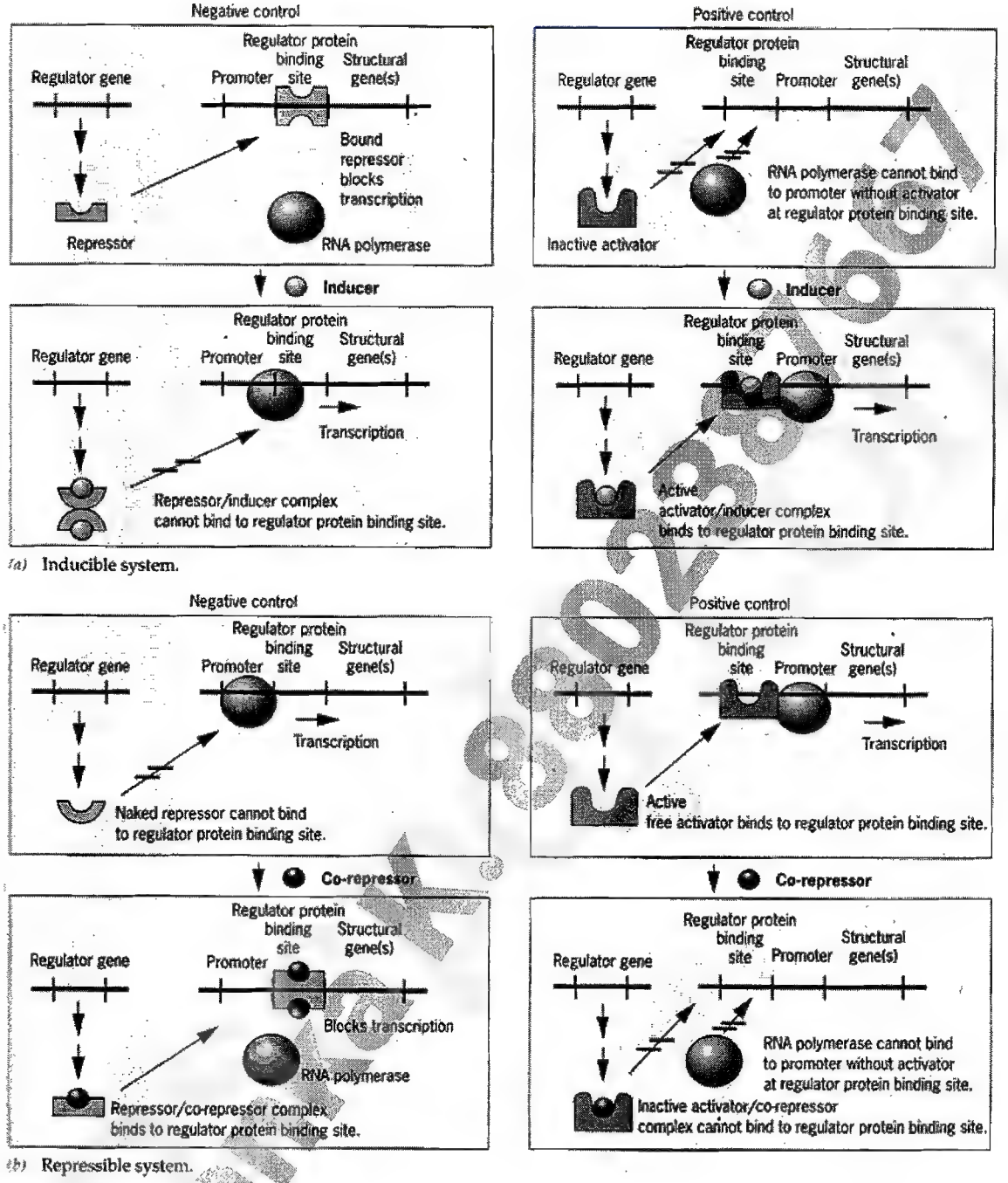


Figure 2: The four broad mechanisms of gene regulation

The strategies of gene regulation in bacterial cells

The mechanisms of gene regulation were first investigated in bacterial cells, in which the availability of mutants and the ease of laboratory manipulation made it possible to unravel the mechanisms.

In bacteria, gene regulation maintains internal flexibility, turning genes on and off in response to environmental changes. The best characterized mechanisms include the following:

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

1. **Control of Transcription Initiation at Operon Circuits**
2. **Use of Alternate Sigma Factors in Controlling Transcription Initiation**
3. **Two Component Based Signaling & Control of Gene Expression**
4. **Premature Transcription Termination by Attenuation**
5. **Anti sense RNA mediated regulation**
6. **Riboswitches & RNA mediated regulation**

Control of Transcription Initiation at Operon Circuits

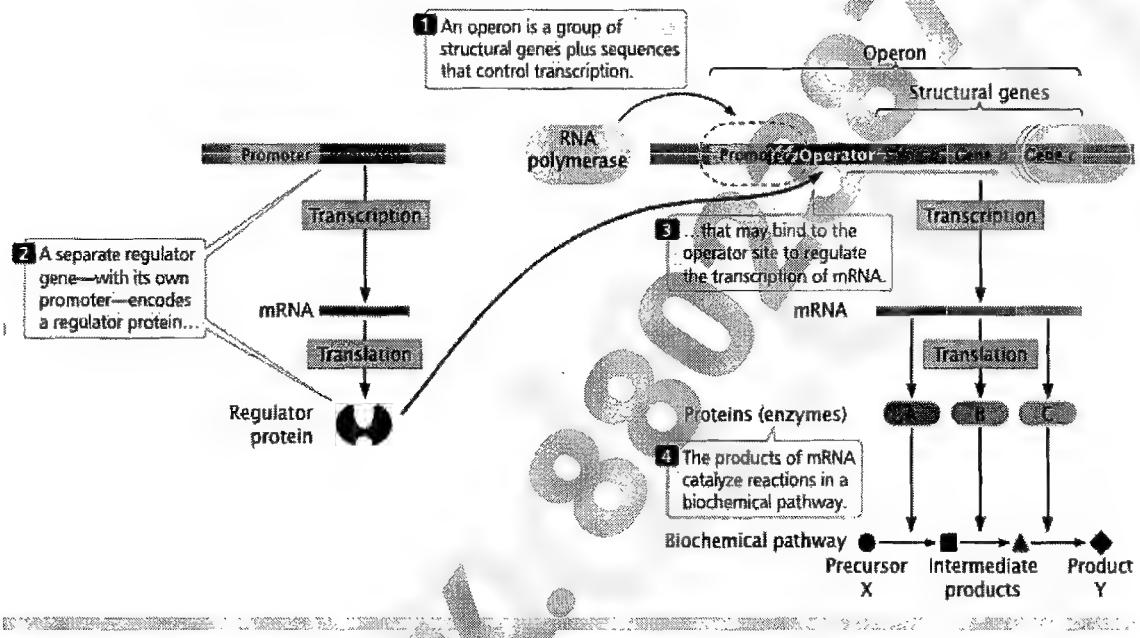


Figure 3: Basic organization of an Operon

This mechanism of gene regulation was first one to be understood (Francois Jacob & Jacques Monod, 1961; Nobel Prize for Medicine in 1965).

Functionally related genes in bacterial cells are frequently clustered together as a single transcriptional unit termed an operon. A typical operon includes several structural genes, a common promoter for the structural genes, and an operator site where the product of a regulator gene binds.

(*Please refer to your class lecture notes on Gene Structure, for a detailed & diagrammatic account of Operon structure*). The organization of a typical operon is illustrated in Figure 3.

The *lac* Operon of *E. coli*

In 1961, Francois Jacob and Jacques Monod described the "operon model" for the genetic control of lactose metabolism in *E. coli*. Operon is the basic unit of transcriptional control in bacteria.

Lactose is one of the major carbohydrates found in milk; it can be metabolized by *E. coli* bacteria that reside in the mammalian gut. Lactose does not easily diffuse across the *E. coli* cell membrane and must be actively transported into the cell by the enzyme permease (Fig. 4, Step 1).

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

To utilize lactose as an energy source, *E. coli* must first break it into glucose and galactose, a reaction catalyzed by the enzyme ß–galactosidase (Fig. 4, Step 2). This enzyme can also convert lactose into allolactose (Fig. 4, Step 3 and 4), a compound that plays an important role in regulating lactose metabolism. A third enzyme, thiogalactoside transacetylase, also is produced by the *lac* operon.

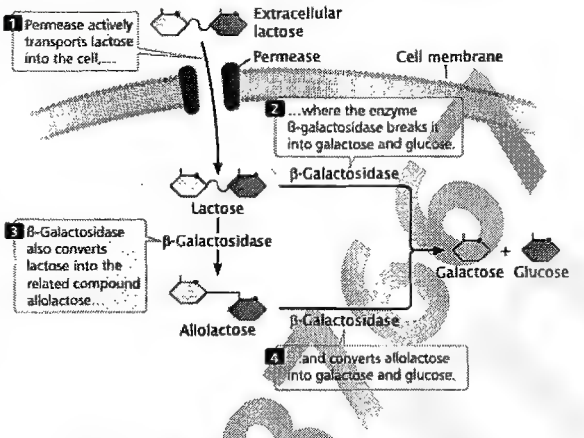


Figure 4: Lactose utilization by *E. coli*

The *lac* operon is an example of a negative inducible operon *but it also has a Positive Control* (as we discuss later). The enzymes ß–galactosidase, permease, and transacetylase are encoded by adjacent structural genes in the *lac* operon of *E. coli*. ß–galactosidase is encoded by the *lacZ* gene, permease by the *lacY* gene, and transacetylase by the *lacA* gene (Fig. 5). When lactose is absent from the medium in which *E. coli* these genes are not transcribed.

If lactose is added to the medium and glucose is absent, the rate of synthesis of all three enzymes simultaneously increases about a thousand fold. This boost in enzyme synthesis results from the transcription of *lacZ*, *lacY*, and *lacA* and exemplifies **coordinate induction**, the simultaneous synthesis of several enzymes, stimulated by a specific molecule, the inducer.

The structure and operation of *lac* operon are described through the step wise diagrams of Figure 5.

Steps in the operation of the *lac* Operon

1. The *lacZ*, *lacY*, and *lacA* genes have a common promoter (*lacP*) and are transcribed together. Upstream of the promoter is a regulator gene, *lacI*, which has its own promoter (P_I).
2. The *lacI* gene is transcribed into a short mRNA that is translated into a repressor. Each repressor consists of **four** identical polypeptides.
3. In the absence of lactose, the repressor binds to the *lac* operator site *lacO*. The operator overlaps the 3′ end of the promoter and the 5′ end of *lacZ*. In 2003, a group of scientists from MIT showed that there are **multiple operators for the *lac* Operon**. The tetrameric repressor binds to two operators together at any given point of time.
4. When the repressor is bound to the operator, the binding of RNA polymerase is blocked, and transcription is prevented.
5. When lactose appears in the medium, the permease that is present transports a small amount of lactose into the cell.
6. Once inside the cell, lactose is converted into allolactose by the enzyme ß-Galactosidase.
7. The allolactose then attaches to the repressor and alters its shape so that the repressor no longer binds to the operator. When the operator site is clear, RNA polymerase can bind and transcribe the structural genes of the *lac* operon. Thus, in the presence of lactose, the

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

repressor is inactivated, the binding of RNA polymerase is no longer blocked, the transcription of *lacZ*, *lacY*, and *lacA* takes place, and the *lac* enzymes are produced.

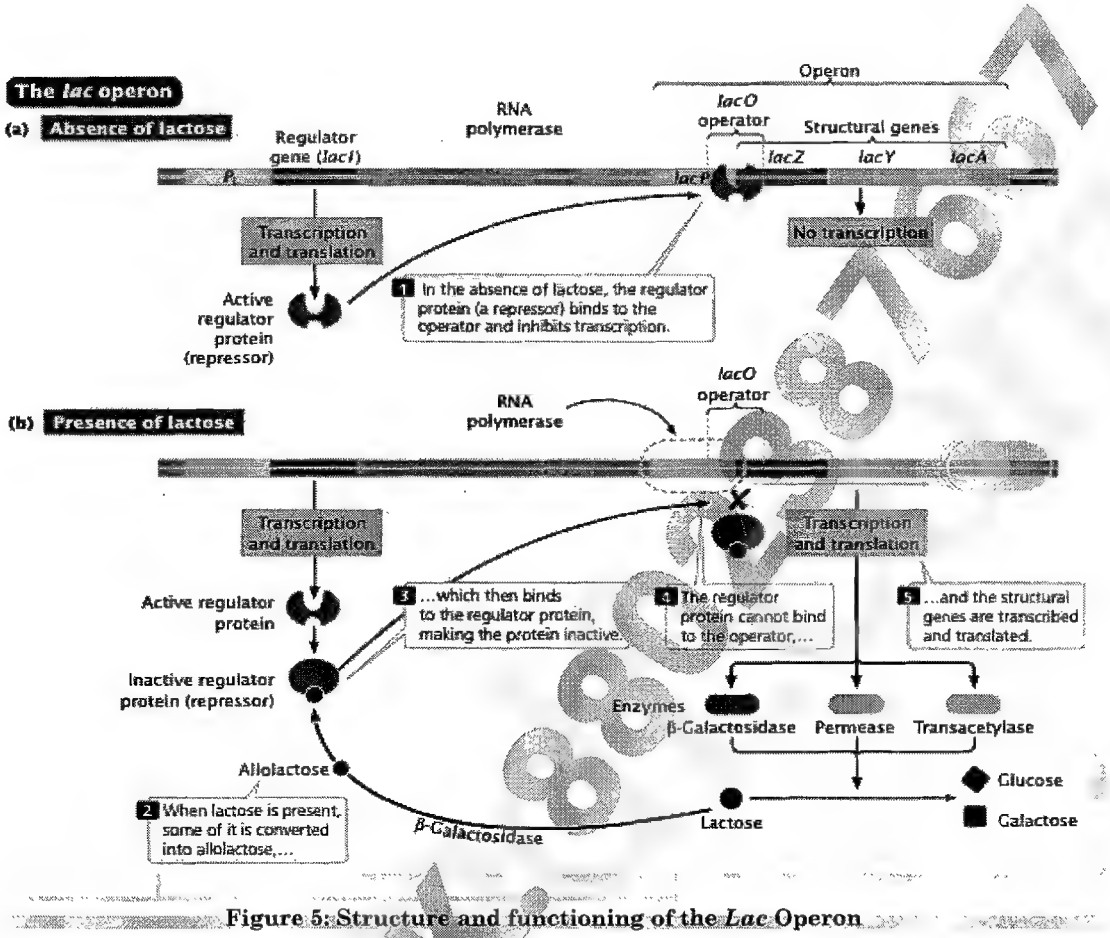


Figure 5: Structure and functioning of the *Lac* Operon

Positive Control of the *lac* Operon

E. coli and many other bacteria will metabolize glucose preferentially in the presence of lactose and other sugars because glucose enters glycolysis without further modification and therefore requires less energy to metabolize than do other sugars. When glucose is available, genes that participate in the metabolism of other sugars are repressed, in a phenomenon known as catabolite repression. For example, the efficient transcription of the *lac* operon takes place only if lactose is present and glucose is absent (Figure 6).

Catabolite repression results from positive control in response to glucose.

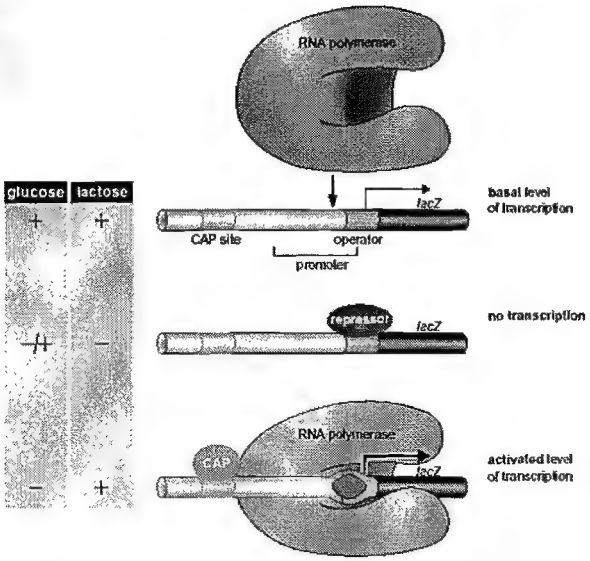


Figure 6: Sugar utilization pattern by *E. coli*

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Positive control is accomplished through the binding of a dimeric protein called the catabolite activator protein (CAP) to a site that is about 22 nucleotides long and is located within or slightly upstream of the promoter of the *lac* genes. RNA polymerase does not bind efficiently to the *lac* promoter unless CAP is first bound to the DNA.

Before CAP can bind to DNA, it must form a complex with adenosine-3′, 5′-cyclic monophosphate (cyclic AMP, or cAMP), which is important in cellular signaling processes in both bacterial and eukaryotic cells. In *E. coli*, the concentration of cAMP is inversely proportional to the level of available glucose. A high concentration of glucose within the cell lowers the amount of cAMP (Fig. 8), and so little cAMP–CAP complex is available to bind to the DNA. Subsequently, RNA polymerase has poor affinity for the *lac* promoter, and little transcription of the *lac* operon takes place. Low concentrations of glucose stimulate high levels of cAMP, resulting in increased cAMP–CAP binding to DNA (Fig 7). This increase enhances the binding of RNA polymerase to the promoter and increases transcription of the *lac* genes by some 50-fold.

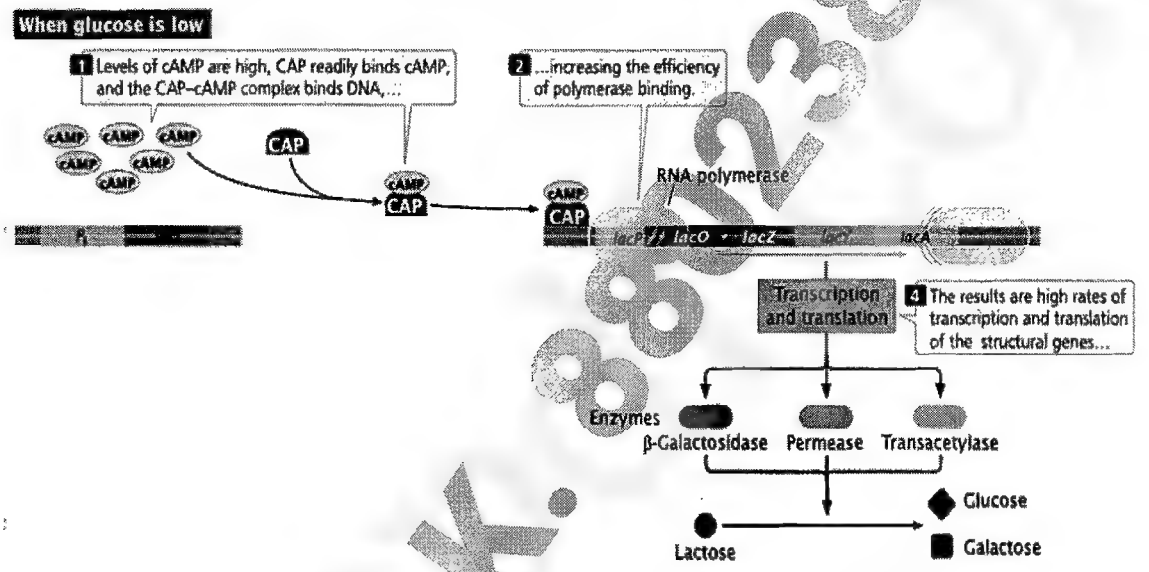


Figure 7: Lac operon functioning, when Glucose level is low.

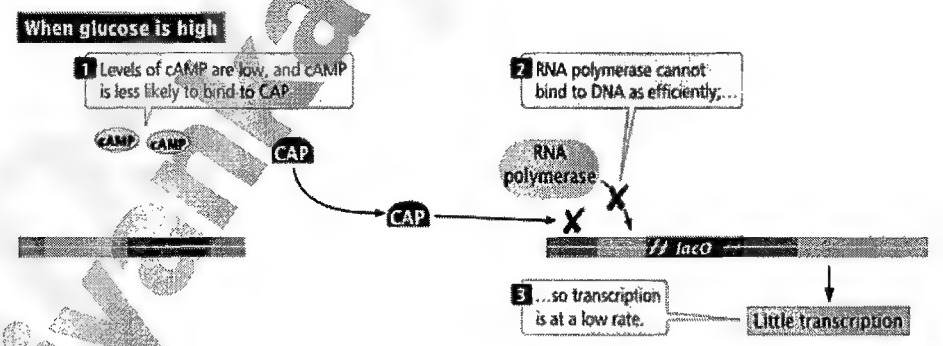


Figure 8: Lac operon functioning, when Glucose level is high.

Use of Alternate Sigma Factors in Controlling Transcription Initiation

The σ subunit of RNA polymerase recognizes the promoter sequences. Most *E. coli* promoters are recognized by RNA polymerase bearing the σ^{70} subunit. *E. coli* encodes several other σ subunits that can replace σ^{70} under certain circumstances and direct the polymerase to alternative promoters. The range of Sigma factors in prokaryotes is tabulated below.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Sigma Factor	Promoters Recognized
σ^{70}	Most genes
σ^{32}	Genes induced by heat shock
σ^{28}	Genes for motility and chemotaxis
σ^{38}	Genes for stationary phase and stress response such as sporulation
σ^{54}	Genes for nitrogen metabolism and other functions

Functioning of alternate Sigma factors can be understood with help of the heat shock σ factor, σ^{32}. Thus, when *E. coli* is subject to heat shock, the amount of this new σ factor increases in the cell, it displaces σ^{70} from most of the RNA polymerases, and directs those enzymes to transcribe genes whose products protect the cell from the effects of heat shock. Level of σ^{32} is increased by two mechanisms – first, its translation is stimulated; and second, the protein is transiently stabilized.

Two Component Based Signaling & Control of Gene Expression

One of the best examples involves the *E. coli* proteins PhoR and PhoB, which regulate transcription in response to the concentration of free phosphate.

1. PhoR is a transmembrane protein, located in the inner (plasma) membrane, whose periplasmic domain (called sensor) binds phosphate and the cytoplasmic domain (called transmitter) is a Kinase.
2. PhoB is the response regulator.

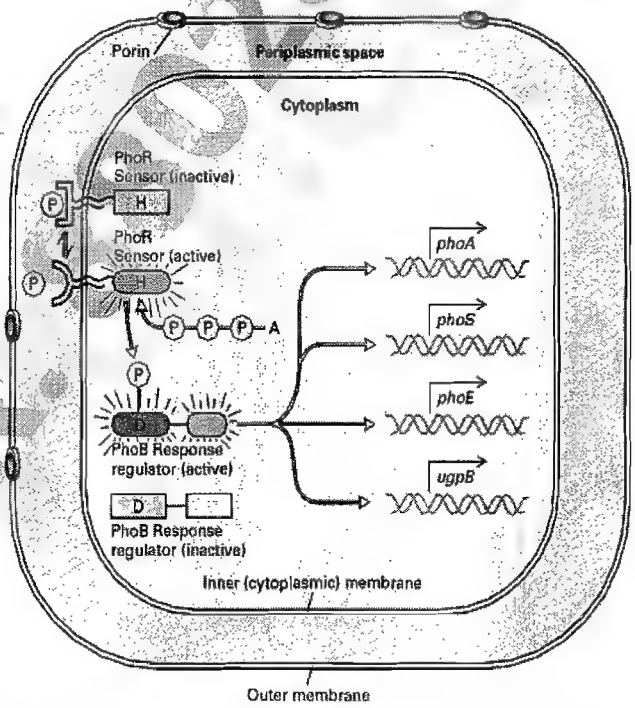


Figure 9: The working of two component system for gene regulation

In response to low phosphate concentrations in the environment and periplasmic space, a phosphate ion dissociates from the periplasmic domain of the sensor protein PhoR causing a conformational change that activates a protein kinase transmitter domain in the cytosolic region of PhoR. Several PhoB proteins can be phosphorylated by one activated PhoR. Phosphorylated PhoB proteins then activate transcription from genes encoding proteins that help the cell to respond to low phosphate, including *phoA, phoS, phoE,* and *ugpB* (Figure 9).

The Premature Termination of Transcription by Attenuation

In **attenuation**, transcription begins at the start site, but termination takes place prematurely, before the RNA polymerase even reaches the structural genes. Attenuation takes place in a number of operons that code for enzymes participating in the biosynthesis of amino acids.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

One of the best-studied examples is found in the *trp* operon of *E. coli* (**Charles Yanofsky et al, 1974**).

The *trp* operon contains a region of 162 nucleotides that is a 5′ UTR. The 5′ UTR (also called a leader) contains four regions: region 1 is complementary to region 2, region 2 is complementary to region 3, and region 3 is complementary to region 4. These complementarities allow the 5′ UTR to fold into two different secondary structures (Fig. 10). Only one of these secondary structures causes attenuation.

One of the secondary structures contains one hairpin produced by the base pairing of regions 1 and 2 and another hairpin produced by the base pairing of regions 3 and 4. A string of uracil nucleotides follows the 3+4 hairpin. Hence the 3+4 hairpin is a terminator, also called an **attenuator**. When cellular levels of tryptophan are high, regions 3 and 4 of the 5′ UTR base pair, to produce the attenuator structure; this base pairing causes transcription to be terminated before the *trp* structural genes can be transcribed.

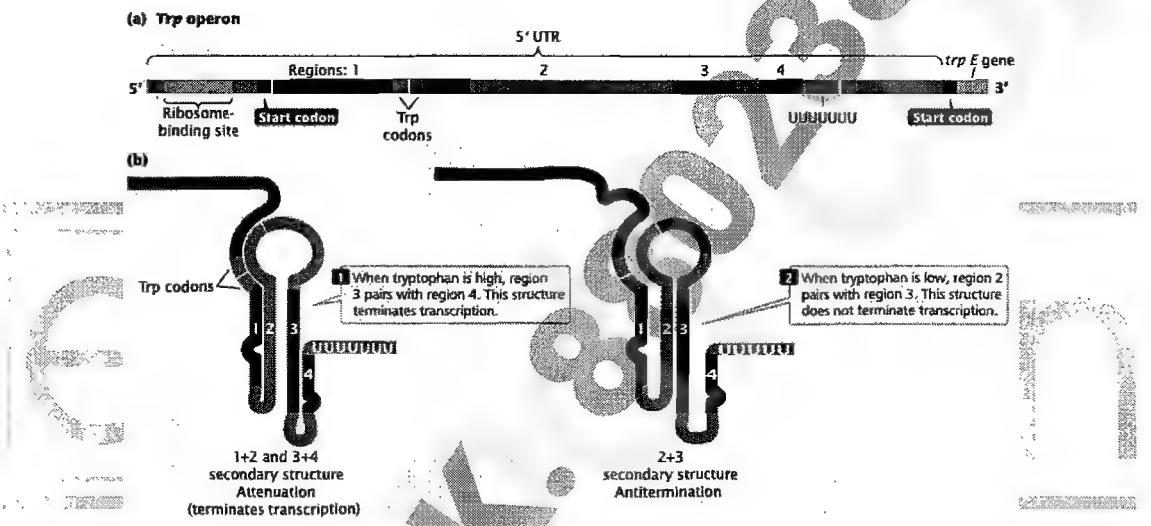


Figure 10: Possible secondary structures in the leader region of *Trp* operon

The alternative secondary structure of the 5′ UTR is produced by the base pairing of regions 2 and 3. This base pairing also produces a hairpin, but this hairpin is not followed by a string of uracil nucleotides; so this structure does not function as a terminator. When cellular levels of tryptophan are low, regions 2 and 3 base pair, and transcription of the *trp* structural genes is not terminated. RNA polymerase continues past the 5′ UTR into the coding section of the structural genes, and the enzymes that synthesize tryptophan are produced. Because it prevents the termination of transcription, the 2+3 structure is called an **antiterminator**.

Why does the 3+4 structure arise when tryptophan is high and the 2+3 structure when tryptophan is low?

Region 1 actually encodes a small protein. Within the coding sequence for this protein are two UGG codons, which specify the amino acid tryptophan; so tryptophan is required for the translation of this 5′ UTR sequence. In prokaryotic cells, transcription and translation are coupled: while transcription is taking place at the 3′ end of the mRNA, translation is initiated at the 5′ end.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Transcription when tryptophan levels are high: (Fig. 11) RNA polymerase begins transcribing the DNA, producing region 1 of the 5′ UTR followed by translation. Meanwhile, RNA polymerase is transcribing region 2. Region 2 is complementary to region 1 but, because the ribosome is translating region 1, the nucleotides in regions 1 and 2 cannot base pair. As RNA polymerase begins to transcribe region 3, the ribosome is continuing to translate region 1. When the ribosome reaches the two UGG tryptophan codons, it doesn't slow or stall, because tryptophan is abundant and tRNAs charged with tryptophan are readily available. Because tryptophan is abundant, translation can keep up with transcription. Even before region 3 is transcribed, the ribosome leaves the segment 1 and partially covers segment 2. Thus, segment 2 does not base pair with segment 3. As a result 3+4 hairpin can also form, as region 4 is transcribed. 3+4 hairpin actually functions as a terminator. Hence transcription is ended even before the structural genes.

Transcription when tryptophan levels are low: (Fig. 11) Once again, RNA polymerase begins transcribing region 1 of the 5. → ribosome also starts translation → but, translation stalls in region 1 itself because tryptophan is scarce → hence region 2 cannot form a base pair loop with region 1 → region 2 is free now → 2+3 loop forms → hence 3+4 loop does not form → termination does not occur.

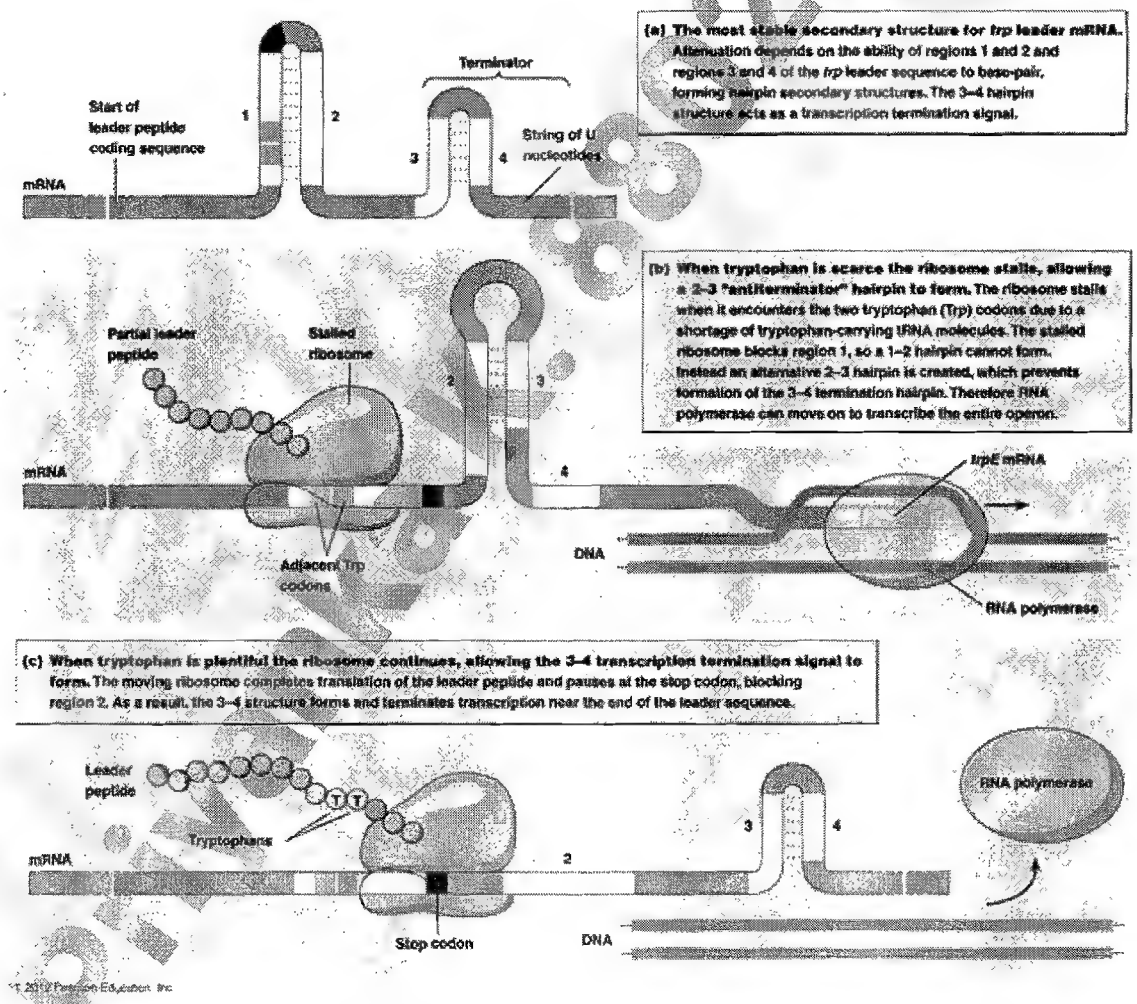


Figure 9: Attentuation of the *Trp* Operon

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Why does attenuation take place?

1. Repression by regulatory proteins is never complete; some transcription is initiated even when the *trp* repressor is active. Attenuation can further reduce transcription. Together the two processes are capable of reducing transcription of the *trp* operon more than 600-fold. Both mechanisms provide *E. coli* with a much finer degree of control over tryptophan synthesis than either could achieve alone.

2. Another reason for the dual control is that attenuation and repression respond to different signals: repression responds to the cellular levels of tryptophan, whereas attenuation responds to the number of tRNAs charged with tryptophan.

3. Finally, the *trp* repressor affects several operons other than the *trp* operon. It's possible that, at an earlier stage in the evolution of *E. coli*, the *trp* operon was controlled only by attenuation. The *trp* repressor may have evolved primarily to control the other operons and only incidentally affects the *trp* operon.

Antisense RNA in Gene Regulation

Several examples of RNA regulators have been discovered in the recent years. Some small RNA molecules are complementary to particular sequences on mRNAs and are called **antisense RNA**. They control gene expression by binding to sequences on mRNA and inhibiting translation.

One example is shown in Figure 12.

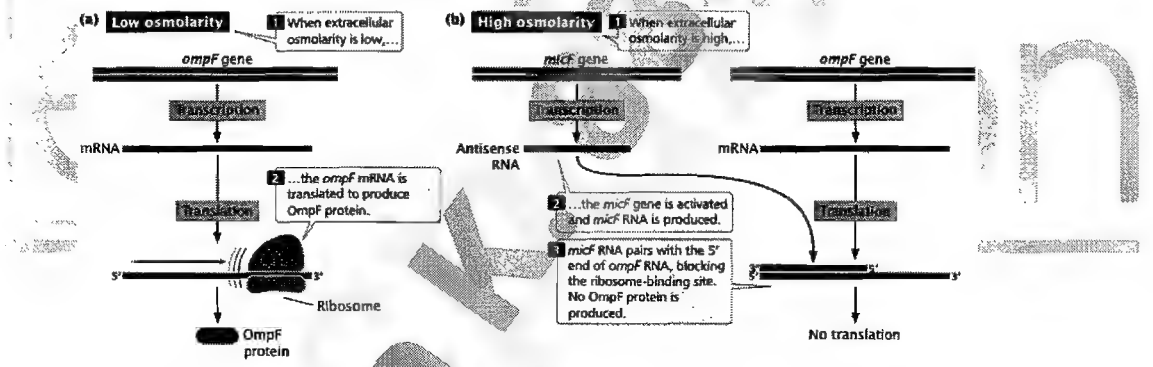


Figure 12: Regulation of gene expression with help of Anti-sense RNA

A number of examples of antisense RNA controlling gene expression have now been identified in bacteria and bacteriophages.

Riboswitches and RNA-Mediated Repression

Some mRNA molecules contain regulatory sequences called **riboswitches**, where molecules may bind and affect gene expression by influencing the formation of secondary structures in the mRNA. In some mRNA molecules, a small regulatory molecule binds to the riboswitch and stabilizes a terminator, which causes premature termination of transcription. In other cases, the binding of a regulatory molecule stabilizes a secondary structure that masks the ribosome-binding site, preventing the initiation of translation. When not bound by the regulatory molecule, the riboswitch assumes an alternative structure that eliminates the premature terminator or makes the ribosome-binding site available.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

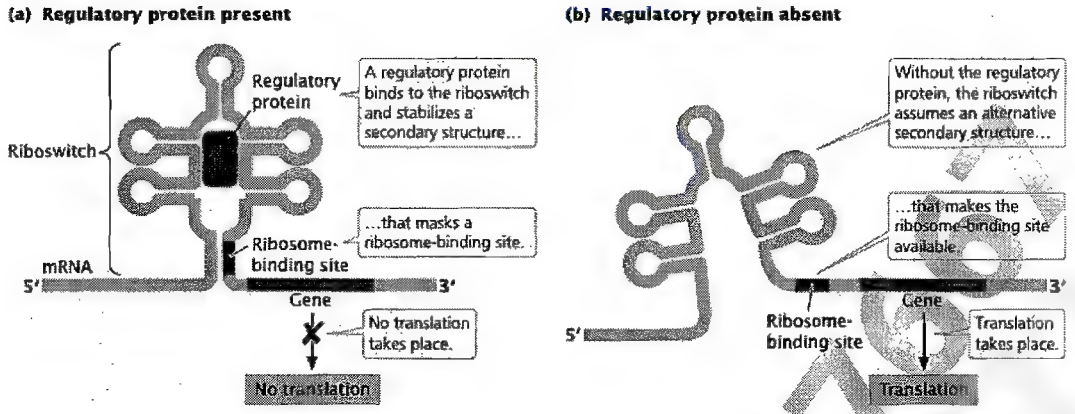


Figure 13: Riboswitch: Structure and Function

An example of a riboswitch is seen in bacterial genes that code for enzymes having roles in the synthesis of vitamin B12. The genes for these enzymes are transcribed into an mRNA molecule with a riboswitch. When the activated form of vitamin B12—called coenzyme B12—is present, it binds to the riboswitch, and the mRNA folds into a secondary structure that masks the ribosome-binding site, and so no translation of the mRNA takes place.

For some riboswitches, the regulatory molecule acts as an inducer by causing a secondary structure that allows transcription or translation to take place. Riboswitches have been observed in eubacteria, archaea, and eukaryotes.

Some other metabolites which bind to the riboswitches are the purines adenine and guanine, the amino acids glycine and lysine, flavin mononucleotide (the prosthetic group of NADH dehydrogenase) etc.

Another type of gene control is carried out by mRNA molecules called ribozymes (Fig. 14), which possess catalytic activity. Termed RNA-mediated repression, this type of control has been demonstrated in the *glmS* gene of the bacterium *Bacillus subtilis*. Transcription of this gene produces an mRNA molecule that codes for the enzyme glutamine-fructose-6-phosphate amidotransferase, which helps synthesize a small sugar called glucosamine-6-phosphate (GlcN6P). Within the mRNA are about 75 nucleotides that act as a ribozyme. When GlcN6P is absent, the *glmS* gene is transcribed and translated to produce the enzyme, which synthesizes more GlcN6P. However, when sufficient GlcN6P is present, it binds to the ribozyme part of the mRNA, which then induces self-cleavage of the mRNA and prevents its translation.

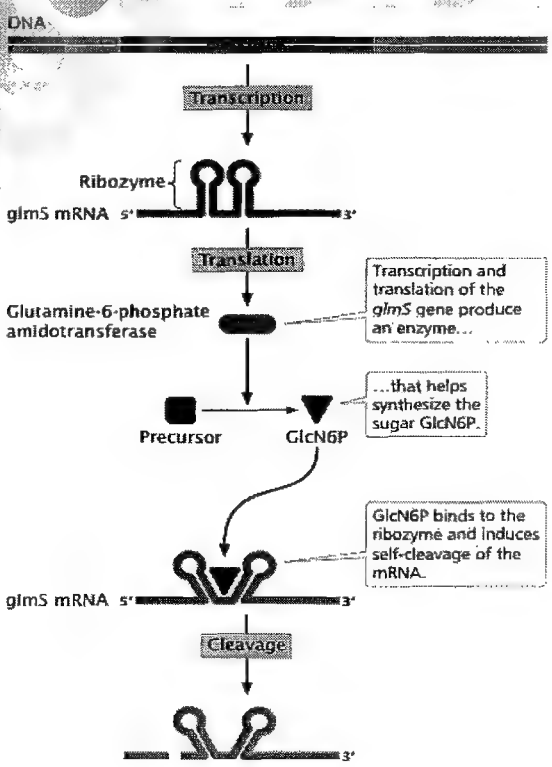


Figure 14: Gene Expression control by Ribozymes

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 8: Regulation of eukaryotic gene expression

Prokaryotes, in general, have a simpler mechanism of gene regulation as compared to the eukaryotes.

- First, most eukaryotic genes are not organized into operons – instead, each structural gene typically has its own promoter and is transcribed separately (*C. elegans* is a notable exception to this generalization).
- Second, chromatin structure affects gene expression in eukaryotic cells; DNA must unwind from the histone proteins before transcription can take place.
- Third, although both repressors and activators function in eukaryotic and bacterial gene regulation, activators seem to be more common in eukaryotic cells.
- Finally, the presence of the nuclear membrane in eukaryotic cells separates transcription and translation in time and space. Therefore, the regulation of gene expression in eukaryotic cells is characterized by a greater of mechanisms that act at different points in the transfer of information from DNA to protein.

A Comparison of Bacterial and Eukaryotic Gene Control

1. Much of gene regulation in both bacterial and eukaryotic cells is accomplished through regulatory proteins that bind to specific sequences in DNA.
2. Both negative control and positive control are found in bacterial and eukaryotic cells.
3. Complex biochemical and developmental events in bacterial and eukaryotic cells may require a cascade of gene regulation, in which the activation of one set of genes stimulates the activation of another set.
4. Most gene regulation in bacterial cells is at the level of transcription (although it does exist at other levels). Gene regulation in eukaryotic cells often takes place at multiple levels, including chromatin structure, transcription, mRNA processing, and RNA stability.
5. In bacterial cells, genes are often clustered in operons and are coordinately expressed by transcription into a single mRNA molecule. In contrast, each eukaryotic gene typically has its own promoter and is transcribed independently. Coordinate regulation in eukaryotic cells takes place through common response elements, present in the promoters and enhancers of the genes. Different genes that have the same response element in common are influenced by the same regulatory protein.
6. Chromatin structure plays a role in eukaryotic (but not bacterial) gene regulation. In general, condensed chromatin represses gene expression; chromatin structure must be altered before transcription can take place. Acetylation of the histone proteins, which may be influenced by the degree of DNA methylation, appears to be important in bringing about these changes in chromatin structure.
7. Transcription is a relatively simple process in bacterial cells, and regulatory proteins function by blocking or stimulating the binding of RNA polymerase to DNA. Eukaryotic transcription requires complex machinery that includes RNA polymerase, general

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

transcription factors, and transcriptional activators, which allows transcription to be influenced by multiple factors.

8. Some eukaryotic transcriptional activator proteins function at a distance from the gene by binding to enhancers, causing the formation of a loop in the DNA, which brings the promoter and enhancer into close proximity. Some distant-acting sequences analogous to enhancers have been described in bacterial cells, but they appear to be less common.

9. The greater time lag between transcription and translation in eukaryotic cells than in bacterial cells allows mRNA stability and mRNA processing to play larger roles in eukaryotic gene regulation.

10. RNA molecules act as regulators of gene expression in both bacterial and eukaryotic systems.

Strategies of Eukaryotic Gene Control

Eukaryotic gene regulation is less well understood than bacterial regulation, partly owing to the larger genomes in eukaryotes, their greater sequence complexity, and the difficulty of isolating and manipulating mutations that can be used in the study of gene regulation. Nevertheless, great advances in our understanding of the regulation of eukaryotic genes have been made in recent years. There are several broad strategies for regulation of eukaryotic gene expression. They are individually described below.

Control at the level of the genome

The well studied examples are outlined below.

1. **Gene amplification,** the selective replication of certain genes. Gene amplification can be regarded as an example of genomic control-that is, a regulatory change in the makeup or structural organization of the genome. One of the best-studied examples of gene amplification involves the ribosomal RNA genes in *Xenopus laevis*. The haploid genome of *Xenopus* normally contains about 500 copies of the genes that code for 5.8S, 18S, and 28S rRNA. During oogenesis (development of the egg prior to fertilization), the DNA of this entire set of genes is selectively replicated about 4000-fold, so that the mature oocyte contains about 2 million copies of the genes for rRNA. Apparently, this level of amplification is necessary to accommodate the enormous amount of ribosome biosynthesis that must take place during oogenesis, which in turn is required to sustain the high rate of protein synthesis needed for early embryonic development. The extra rRNA gene copies created by gene amplification are present in extrachromosomal circles of DNA distributed among hundreds of nucleoli that appear in the oocyte nucleus as amplification progresses.

   Other example of gene amplification is the creation of **polytene chromosomes**. To increase cell volume, some specialized cells undergo repeated rounds of DNA replication without cell division (endomitosis), forming a giant polytene chromosome.

2. **Deletion:** Some cells delete genes whose products are not required. An extreme example of **gene deletion** (also called ***DNA diminution***) occurs in mammalian red blood cells, which discard their nuclei entirely after adequate amounts of hemoglobin mRNA have been made. Another example occurs in a group of tiny crustaceans known as **copepods** where during the embryonic development of copepods, the heterochromatic regions of their chromosomes are excised and discarded from all cells except those destined to become gametes.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

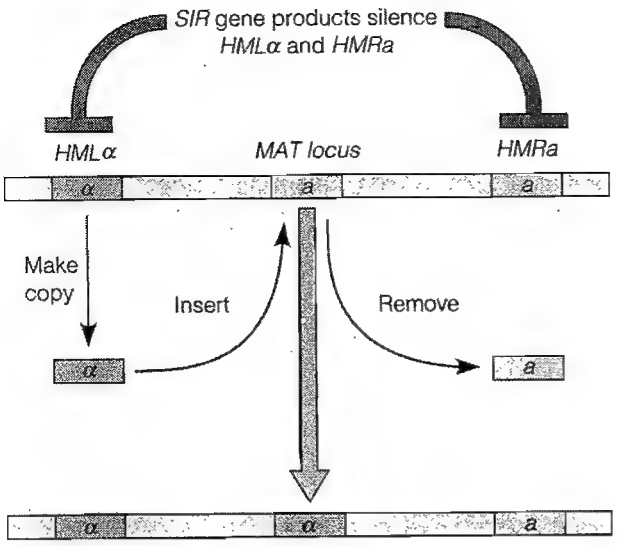


Figure 1: The Cassette Mechanism of the Yeast Mating-Type Switch. Chromosome 3 of *Saccharomyces cerevisiae* contains three copies of the mating-type information. The *HMLa* and *HMRα* loci contain complete copies of the *a* and *α* forms of the gene, but the cell's actual mating type is determined by the allele present at the *MAT* locus. When a cell switches mating types, the *a* or *α* DNA at the *MAT* locus is removed and replaced by a DNA copy of the alternative mating-type DNA. As an example, this figure illustrates a switch in mating type from *a* to *α*.

3. Genomic rearrangements:

a. Yeast Mating-Type Rearrangements. In the yeast *Saccharomyces cerevisiae,* mating occurs when haploid cells of two different mating types, called *a* and *α,* fuse together to form a diploid cell. All haploid cells carry both alleles for mating type; however, a cell's actual mating phenotype depends on which of the two alleles, *a* or *α,* is present at a special site in the genome called the ***MAT* locus.** Cells frequently switch mating type, presumably as a means of maximizing opportunities for mating. They do so by moving the alternative allele into the *MAT* locus. This process of DNA rearrangement is called the **cassette mechanism,** because the mating-type locus is like a tape deck into which either the *a* or the *α* "cassette" (allele) can be inserted and "played" (transcribed). Figure 1 describes the yeast cassette mechanism.

Antibody Gene Rearrangements. A somewhat different type of DNA rearrangement is used by lymphocytes of the vertebrate immune system for producing antibody molecules. Antibodies are proteins composed of two kinds of polypeptide subunits, called *heavy chains* and *light chains.* The rearrangement process involves four kinds of DNA sequences, called *V,* J, D, and C *segments.* The C segment codes for a heavy or light chain *constant region* whose amino acid sequence is the same among different antibodies; the V, J, and D segments together code for *variable regions* that differ among antibodies and give each one the ability to recognize and bind to a specific type of foreign molecule.

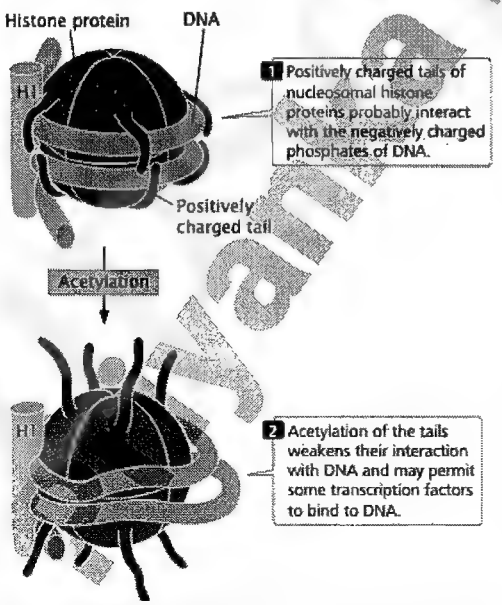


Figure 3: Effect of Histone Acetylation on chromatin packaging.

Control of Gene Access

This is the major line of gene control in eukaryotic cells in which before transcription, chromatin structure changes, and the DNA becomes more accessible to the transcriptional machinery.

The chromatin structure represses gene expression.

The methods by which chromatin structure is altered are as follows.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

1. **Histone acetylation & deacetylation:** Histones in the octamer core of the nucleosome have two domains: (1) a globular domain that associates with other histones and the DNA and (2) a positively charged tail domain that probably interacts with the negatively charged phosphates on the backbone of DNA.

   Acetyl groups are added to histone proteins by **histone acteyltransferase (HAT)** enzymes; the acetyl groups destabilize the nucleosome structure, perhaps by neutralizing the positive charges on the histone tails and allowing the DNA to separate from the histones. Example of HAT - *Tetrahymena* protein called **p55,** mammalian protein called **p300/CBP**.

   Other enzymes called **histone deacetylases (HDACs)** strip acetyl groups from histones and restore chromatin condense structure.

   In addition to acetylation, **histone** proteins may be modified by the addition of phosphates (phosphorylation) and methyl groups (methylation). Both types of modification alter **chromatin** structure and potentially affect **transcription.**

2. **DNA methylation:** Another change in chromatin structure associated with transcription is the methylation of cytosine bases, which yields 5–methylcytosine. Heavily methylated DNA is associated with the gene repressors called **Methylated Cytosine Binding Proteins (MeCP)**. This binding causes repression of transcription in vertebrates and plants, whereas transcriptionally active DNA is usually unmethylated in these organisms.

   DNA methylation is most common on cytosine bases adjacent to guanine nucleotides on the same strand (CpG); so two methylated cytosines sit diagonally across from each other on opposing strands:

   . . . GC . . .

   . . . CG . . .

   DNA regions with many CpG sequences are called **CpG islands** and are found near transcription start sites of nearly 60% of the eukaryotic genes. While genes are not being transcribed, these CpG islands are often methylated, but the methyl groups are removed before the initiation of transcription.

   X chromosome Inactivation in Mammals by CpG Methylation

   CpG methylation is also associated with long-term gene repression, such as on the inactivated X chromosome of female mammals.

   X inactivation a special form of imprinting that leads to total inactivation of one of the X chromosomes in a female mammalian cell. It occurs because females have two X chromosomes whereas males have only one. If both of the female X chromosomes were active then proteins coded by genes on the X chromosome might be synthesized at twice the rate in females compared with males. To avoid this undesirable state of affairs, one of the female X chromosomes is silenced and is seen in the nucleus as a condensed structure called the **Barr body** , which is comprised entirely of heterochromatin. Silencing occurs early in embryo development and is controlled by the X inactivation center, a discrete region present on each X chromosome.

   In a cell undergoing X inactivation, the inactivation center on one of the X chromosomes initiates the formation of heterochromatin, which spreads out from the nucleation point until

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

the entire chromosome is affected, with the exception of a few short segments containing small clusters of genes that remain active. The process dependent upon, each of the following:

- A gene called *Xist*, located in the inactivation center, which is transcribed into a 25-kb non-coding RNA, copies of which coat the chromosome as heterochromatin is formed;
- Replacement of histone H2A, one of the members of the core octamer of the nucleosome with a special histone, macroH2A1;
- Deacetylation of histone H4, as usually occurs in heterochromatin;
- Hypermethylation of certain DNA sequences, although this appears to occur after the inactive state has been set up.

X inactivation is heritable and is displayed by all cells descended from the initial one within which the inactivation took place.

3. **Cross Talk between Histone Deacetyltion and DNA Methylation:** Evidence indicates that ***methylation appears to attract deacetylases***, which remove acetyl groups from the histone tails, stabilizing the nucleosome structure and repressing transcription. Demethylation of DNA would allow acetyltransferases to add acetyl groups, disrupting nucleosome structure and permitting transcription. Now it is known that methyl-CpG-binding proteins **(MeCPs)** are components of both the Sin3 and NuRD histone deacetylase complexes.
4. **Chromatin remodeling:** Some transcription factors and other regulatory proteins alter chromatin structure without affecting the histones directly. These **chromatin-remodeling complexes** bind directly to particular sites on DNA and reposition the nucleosomes, allowing transcription factors to bind to promoters and initiate transcription. The proteins responsible for nucleosome remodeling work together in large complexes. One of these is **Swi/Snf**, made up of at least 11 proteins, which is present in many eukaryotes.
5. **Other Chemical modifications of Histones:**

- If histone tails are methylated, it blocks the site of action of HATs. Thus chromosome remains in the chromatin state solenoid structure and transcription does not take place. For transcription to take place now, the following steps must be undertaken:

  Removal of methyl groups by Histone demethylases → Action of HATs → Chromatin opening → Transcription.

- A number of recent reports indicate that the effects of methylation of a pair of lysine amino acids at the fourth and ninth positions from the N-terminus of histone H3 are particularly interesting. Methylation of lysine-9 induces chromatin packaging and silences gene expression, but methylation of lysine-4 has the opposite effect and promotes an open chromatin structure. Within the ß-globin functional domain, and probably elsewhere, lysine-4 methylation is closely correlated with acetylation of histone H3, and the two types of modification may work hand in hand to activate regions of chromatin.
- The tails of the core histones also have attachment sites for phosphate groups and for the common ('ubiquitous') protein called ubiquitin . They too can compact the chromatin

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

structure and repress gene expression. Similar effect is also seen when **Small Ubiquitin like Modifiers (SUMOs)** bind to the chromatin complex.

- Phosphorylation of histone H3 and of the linker histone has been associated with formation of metaphase chromosomes.

Our growing awareness of the variety of histone modifications that occur, and of the way in which different modifications work together, has led to the suggestion that there is a **Histone Code** (The hypothesis that the pattern of chemical modification on histone proteins influences various cellular activities) by which the pattern of chemical modifications specifies which regions of the genome are expressed at a particular time (Strahl and Allis, 2000; Jenuwein and Allis, 2001, 2005).

Transcriptional Control in Eukaryotic Cells

General transcription factors and RNA polymerase assemble into a *basal transcription apparatus*, which binds to a *core promoter* located immediately upstream of a gene. The basal transcription apparatus is capable of minimal levels of transcription; transcriptional activator proteins are required to bring about normal levels of transcription. These proteins bind to a **regulatory promoter** and to ***enhancers***.

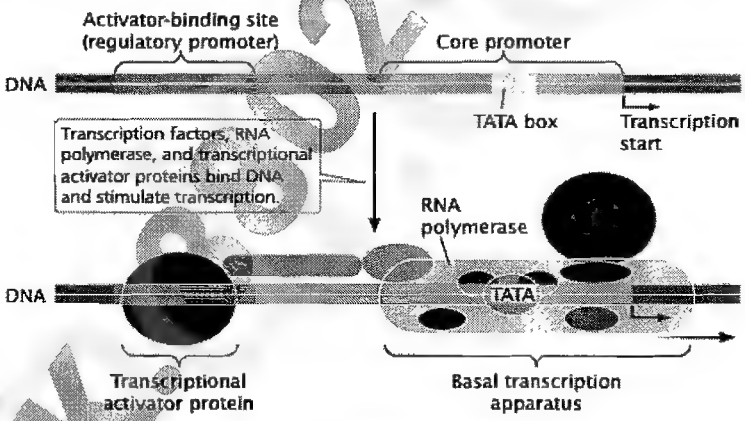


Figure 4: Action of Transcription Activator Proteins

Transcriptional activator proteins stimulate and stabilize the basal transcription apparatus at the core promoter. The activators may interact directly with the basal transcription apparatus or indirectly through protein **coactivators**. Some activators and coactivators, as well as the general transcription factors, also have acteyltransferase activity and so further stimulate transcription by altering chromatin structure.

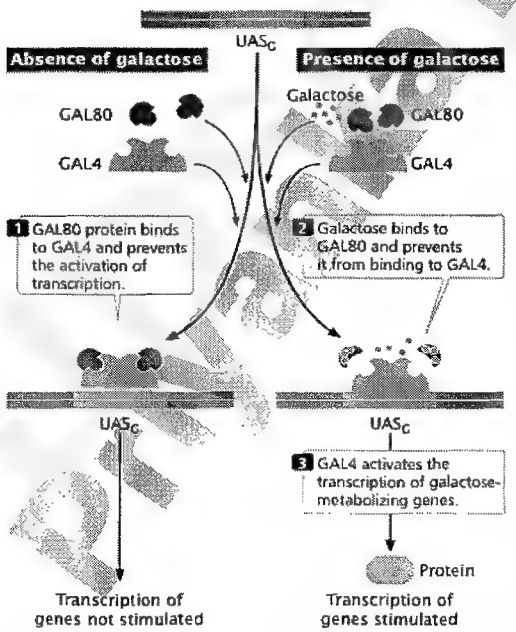


Figure 5: Action of GAL4 transcriptional activator protein

Transcriptional activator proteins have two distinct functions. First, they are capable of binding DNA at a specific base sequence, usually a consensus sequence in a regulatory promoter or enhancer. A second function is the ability to interact with other components of the transcriptional apparatus and influence the rate of transcription.

GAL4 is a transcriptional activator protein that regulates the transcription of several yeast genes in galactose metabolism. When bound to UASG, GAL4 activates the transcription of yeast genes needed for metabolizing galactose. A particular

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

region of GAL4 binds another protein called GAL80, which regulates the activity of GAL4 in the presence of galactose. When galactose is absent, GAL80 binds to GAL4 (two molecules of GAL80 bind to each molecule of GAL4), preventing GAL4 from activating transcription.

When galactose is present, however, it binds to GAL80, causing a conformational change in the protein so that it can no longer bind GAL4. The GAL4 protein is then available to activate the transcription of the genes whose products metabolize galactose.

Some regulatory proteins in eukaryotic cells act as **repressors**, inhibiting transcription. These repressors may bind to sequences in the regulatory promoter or to distant sequences called ***silencers***, which, like enhancers, are position and orientation independent. Unlike repressors in bacteria, most eukaryotic repressors do not directly block RNA polymerase.

Enhancers are capable of affecting transcription at distant promoters. A typical enhancer is some 500 bp in length and contains 10 binding sites for proteins that regulate transcription. Mechanism - activator proteins bind to the enhancer and cause the DNA between the enhancer and the promoter to loop out, bringing the promoter and enhancer close to one another, and so the transcriptional activator proteins are able to directly interact with the basal transcription apparatus at the core promoter. Some enhancers may take help of a **multi-protein mediator complex**.

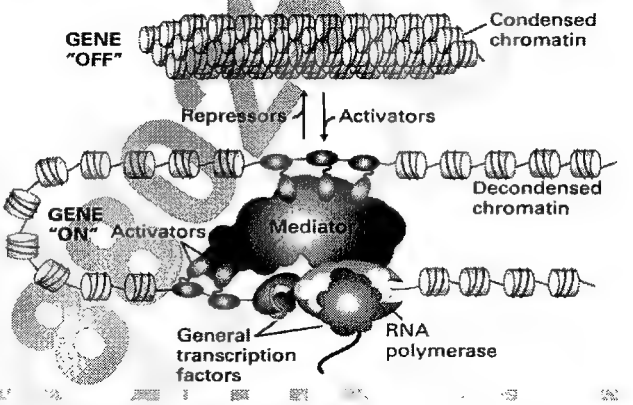


Figure 6: The effect of Enhancer elements on Gene Expression

Coordinated Gene Regulation

Although eukaryotic cells do not possess operons, several eukaryotic genes may be activated by the same stimulus. For example, many eukaryotic cells respond to extreme heat and other stresses by producing **heat-shock proteins** that help to prevent damage from such stressing agents. During times of environmental stress, the transcription of all the heat-shock genes is greatly elevated.

Genes that are coordinately expressed in eukaryotic cells are able to respond to the same stimulus because they have regulatory sequences in common in their promoters or enhancers. For example, different eukaryotic heat-shock genes possess a common regulatory element upstream of their start sites. A transcriptional activator protein binds to this regulatory element during stress and elevates transcription. Such common DNA regulatory sequences are called **response elements**; they typically contain short consensus sequences.

A few response elements found in eukaryotic cells		
Response element	**Responds to**	**Consensus sequence**
Heat-shock element	Heat and other stress	CNNGAANNTCCNNG
Glucocorticoid response element	Glucocorticoids	TGGTACAAATGTTCT
Phorbol ester response element	Phorbal esters	TGACTCA
Serum response element	Serum	CCATATTAGG

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Gene Control through Messenger RNA Processing

A. Alternative splicing

The primary transcripts of some genes can follow two or more **alternative splicing** pathways, enabling a single transcript to be processed into related but different mRNAs and hence to direct synthesis of a range of proteins. Only three examples being known in *Saccharomyces cerevisiae*, but in higher eukaryotes it is much more prevalent.

At least 38% of the genes in the human genome undergo alternative splicing (Human Genome Report, 2005): the principle 'one gene, one protein', biological dogma since the 1940s, has been completely overthrown.

Alternative splicing is now looked on as a crucial innovation in the genome expression pathway.

Important examples

- The determination of sex in *Drosophila* is determined by an alternative splicing cascade.
- The human *slo* gene codes for a membrane protein that regulates the entry and exit of potassium ions into and out of cells (Graveley, 2001). The gene has 35 exons, eight of which are involved in alternative splicing events, leading to over 500 distinct mRNAs, each specifying a membrane protein with slightly different functional properties. The human *slo* genes are active in the inner ear and determine the auditory properties of the hair cells on the basilar membrane of the cochlea. Alternative splicing of *slo* genes in cochlear hair cells therefore determines the auditory range of humans.

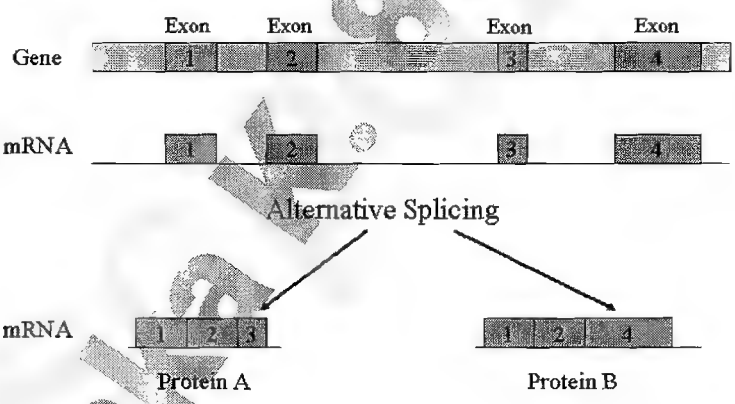


Figure 7: Alternative Splicing

B. mRNA Editing

RNA editing involves alteration of a base by specific enzymes in such a way that it alters the coding properties of the transcript. It is relatively more frequent in eukaryotes than it was earlier believed to be. One of the best-studied examples includes the expression of the mRNA for the Apolipoprotein B [ApoB] of mammalian digestive system.

In mammals, the Apo-B gene is expressed in both hepatocytes (liver cells) and intestinal epithelial cells. However, in liver cells, its product is a 500 kD protein called Apo-B100 whereas in intestine cells its product is a smaller protein called Apo-B48. The Apo-B100 is produced without RNA editing, but the Apo-B48 is synthesized from an mRNA whose sequence has been

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

altered by a specific enzyme. This enzyme changes a codon, CAA, in the middle of the original mRNA to the stop codon UAA, thereby causing early termination of the protein synthesis.

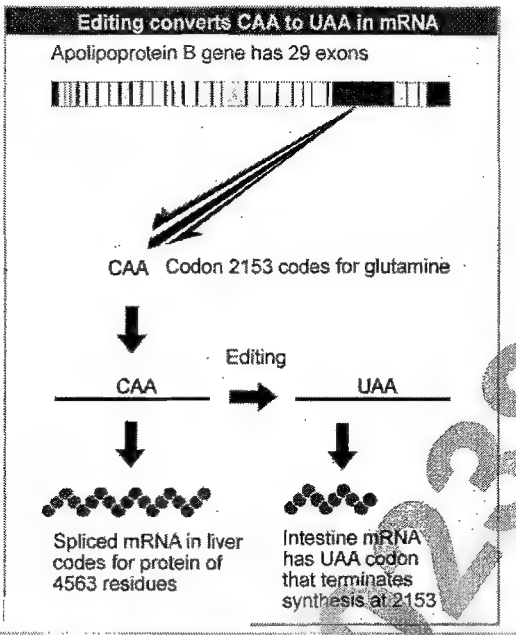


Figure 8: Editing of ApoB mRNA in humans

Other examples of mRNA editing in mammals

Tissue	Target RNA	Change	Comments
Intestine	Apolipoprotein B mRNA	C→U	Converts a glutamine codon to a stop codon
Muscle	α-galactosidase mRNA	U→A	Converts a phenylalanine codon into a tyrosine codon
Testis, tumors	Wilms tumor-1 mRNA	U→C	Converts a leucine codon into a proline codon
Tumors	Neurofibromatosis type-1 mRNA	C→U	Converts an arginine codon into a stop codon
B lymphocytes	Immunoglobulin mRNA	Various	Contributes to the generation of antibody diversity
HIV-infected cells	HIV-1 transcript	G→A, C→U	Involved in regulation of the HIV-1 infection cycle
Brain	Glutamate receptor mRNA	A→inosine	Multiple positions leading to various codon changes

Gene Control through RNA Stability

The PolyA tail on the 3' end of the nascent transcript is now clearly linked to delay the degradation of the mRNA's coding regions by the cytosolic and even nuclear 3'→5' exonucleases, the most abundant class of exonucleases. The mRNA with a longer PolyA tail translates multiple times in the cytosol, while those with a small PolyA tail, such as the **mRNA for Histone-1**,

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

degrade only after a few rounds of translation. This mechanism ensures that those mRNA whose translation products are needed in large quantities are provided with a long PolyA tail.

mRNA stability is sensitive to the extracellular signals too. For example, in-vitro analysis has shown that prolactin treatment of cultured breast tissue causes great increase in the stability of casein mRNA.

RNA induced silencing

Growing evidence indicates that the expression of a number of eukyarotic genes is controlled through RNA interference, also known as RNA silencing and posttranscriptional gene silencing. Two US geneticists Andrew Fire and Craig Mello, who revealed the process of RNA interference (RNAi) in 1998 received Nobel Prize in 2006.

The mechanism of RNA induced silencing is shown below.

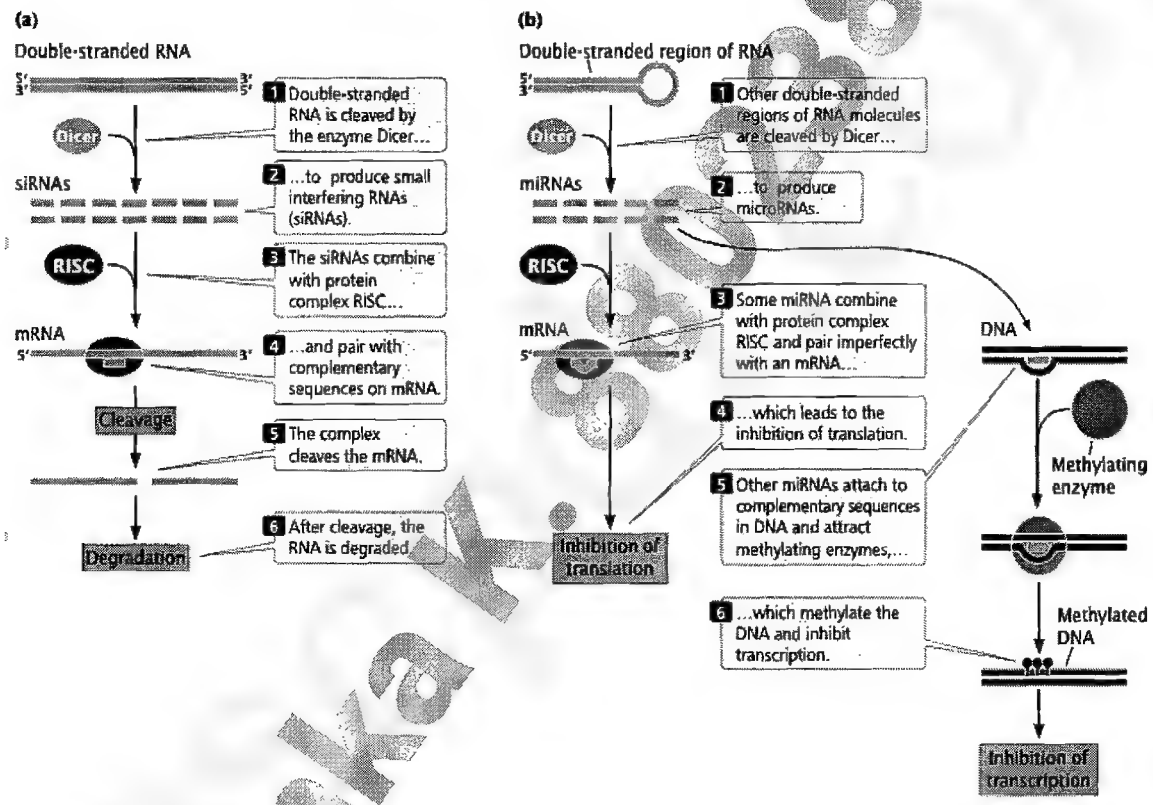


Figure 9: RNA induced gene silencing mechanism

RNA interference appears to be widespread, existing in fungi, plants, and animals. RNAi occurs naturally to regulate gene expression and control viral infections. This technique is also widely used as a powerful tool for artificially regulating gene expression in genetically engineered organisms.

RNA silencing is mediated by very small RNA molecules know as microRNAs (miRNAs) and small interferring RNAs (siRNAs), depending on their origin and mode of action.

These RNA molecules originate from double-stranded RNA, which may arise in several ways: when a single-stranded RNA molecule base pairs with itself to form double-stranded regions of RNA; by the simultaneous transcription of two different RNA molecules that are complementary

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

to one another and that pair to form double-stranded RNA; or by the replication of double-stranded RNA viruses.

In RNA silencing, an enzyme called Dicer cleaves and processes double-stranded RNA to produce siRNAs or miRNAs that are 21 to 25 nucleotides in length and pair with proteins to form an RNA-induced silencing complex (RISC). The RNA component of the RISC then pairs with complementary base sequences of specific mRNA molecules; siRNAs tend to base pair perfectly with the mRNAs, whereas miRNAs form less-than-perfect pairings. In the siRNA case, RISC then cleaves the mRNA near the middle of the bound siRNA. After cleavage, the mRNA is further degraded. In the miRNA case, the RISC inhibits translation of the mRNA. For example, an important gene in flower development in *Arabidopsis thaliana* is *APETALA2*. The expression of this gene is regulated by an miRNA that base pairs with nucleotides in the coding region of *APETALA2* mRNA and inhibits its translation. Some miRNAs serve as guides for the methylation of complementary sequences in DNA, which then inhibits transcription.

Riboswitches and Ribozymes

Same as discussed in Prokaryotic Regulation.

Translational Control

Ribosomes, aminoacyl tRNAs, initiation factors, and elongation factors are all required for the translation of mRNA molecules. The availability of these components affects the rate of translation and therefore influences gene expression.

There are two types of translational control.

1. Specialized translational control (seen in case of *Ferritin* mRNA translation and described in Figure 10) ; and
2. Generalized translational control (seen in case of *Globin* mRNA translation and described in Figure 11)

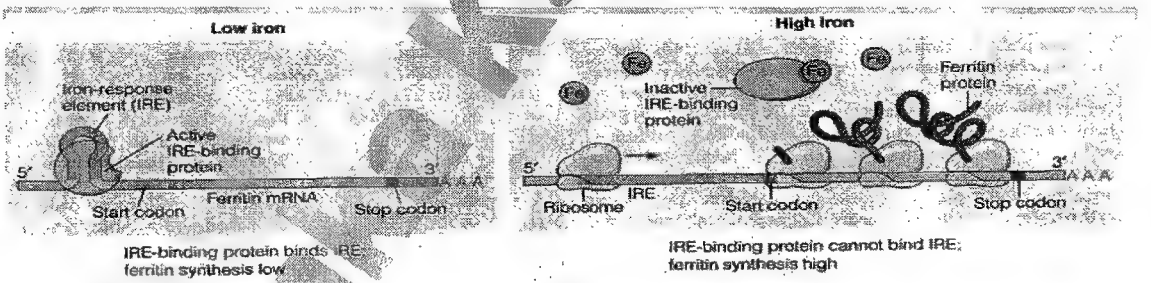


Figure 10: A specific type of translational control is seen with *ferritin*, an iron-storage protein: This protein's synthesis is selectively stimulated in the presence of iron. The key to this selective stimulation lies in the 5' untranslated leader sequence of *ferritin* mRNA, which contains a 28-nucleotide segmentthe iron-response element (IRE)-that forms a hairpin loop required for the stimulation of ferritin synthesis by iron. When the iron concentration is low (left panel), a regulatory protein called the *IRE-binding protein* binds to the IRE sequence in ferritin mRNA, preventing the mRNA from forming an initiation complex with ribosomal subunits. But the IRE-binding protein is an allosteric protein whose activity can be controlled by the binding of iron. When more iron is available (right panel), the protein binds an iron atom and undergoes a conformational change that prevents it from binding to the IRE, thereby allowing the *ferritin* mRNA to be translated. The IREbinding protein is therefore an example of a translational repressor that selectively controls the translation of a particular mRNA. This type of translational control allows cells to respond to specific changes in the environment faster than would be possible by transcriptional control.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

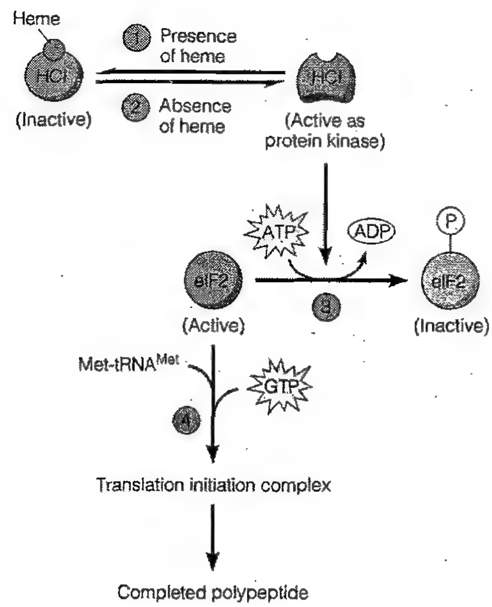


Figure 11: Regulation of Translation by Heme in Developing Erythrocytes. The main function of developing erythrocytes is to synthesize hemoglobin, which consists of four globin polypeptides and a heme prosthetic group. These cells contain the protein HCI (heme-controlled inhibitor), which regulates this synthesis in response to the availability of heme. (1) When heme is present, it binds to HCI inactivating it. (2) When heme is absent, HCI is active. (3) Active HCI functions as a kinase that catalyzes the phosphorylation of e1F2, a key translation initiation factor. Phosphorylated eIF2 is inactive; it cannot combine with methionyl-tRNA and GTP to form the translation initiation complex. Thus, in the absence of heme, translation of all mRNA in the cell is inhibited. The main effect is on globin synthesis, because globin mRNA constitutes most of the cell's mRNA. (4) When heme is present, translation of the mRNA proceeds. Newly made globins combine with heme to form hemoglobin molecules.

Conclusion

Transient changes in genome activity occur predominantly in response to external stimuli. For unicellular organisms, the most important external stimuli relate to nutrient availability, these cells living in variable environments in which the identities and relative amounts of the nutrients change over time. The genomes of unicellular organisms therefore include genes for uptake and utilization of a range of nutrients, and changes in nutrient availability are shadowed by changes in genome activity, so that at any one time only those genes needed to utilize the available nutrients are expressed. Most cells in multicellular organisms live in less variable environments, but an environment whose maintenance requires coordination between the activities of different cells. For these cells, the major external stimuli are therefore hormones, growth factors, and related compounds that convey signals within the organism and stimulate coordinated changes in genome activity.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 9: Gene silencing

Gene silencing is a general term describing epigenetic processes of gene regulation. The term gene silencing is generally used to describe the "switching off" (or a non-expressional state) of a gene by a *mechanism other than genetic modification*, such as mutations, recombination, transpositions etc. Thus, a gene, which would be expressed (turned on) under normal circumstances, is switched off by machinery in the cell.

There are two main levels and types of gene silencing.

1. **Transcriptional gene silencing** is the result of:

   1a. Histone modifications of various types – (Please refer to your notes on Eukaryotic Gene Regulation for details).

   1b. DNA Methylation – (Please refer to your notes on Eukaryotic Gene Regulation for details).

   1c. Action of structural RNA such as **Xist** during Barr Body formation in mammalian females. Xist is an RNA gene on the X chromosome of the placental mammals that acts as major effector of the X inactivation process. The Xist RNA, a large (17 kb in humans) transcript, is expressed on the inactive chromosome and not on the active one. It is processed similarly to mRNAs, through splicing and polyadenylation, however, it remains untranslated. The inactive X is coated with this transcript, which is essential for the inactivation. X lacking Xist will not be inactivated, while duplication of the Xist gene on another chromosome causes inactivation of that chromosome.

   1d. Binding of gene silencing proteins, such as the **Polycomb Group Proteins (PcG Proteins)** in *Drosophila* and several other organisms. Polycomb-group proteins are a family of proteins, which can remodel chromatin in such a way that transcription factors cannot bind to promoter sequences in DNA. Polycomb group proteins play a role in silencing HOX genes through modulation of chromatin structure.

   In *Drosophila*, the Trithorax-group (trxG) and Polycomb-group (PcG) proteins act antagonistically and interact with chromosomal elements, termed **Cellular Memory Modules** (CMMs). Trithorax-group (trxG) proteins maintain the active state of gene expression while the Polycomb-group (PcG) proteins counteract this activation with a repressive function that is stable over many cell generations. Polycomb Gene complex induced silencing involves at least three kinds of multiprotein complex PRC1, PRC2 and PhoRC which work together to carry out their repressive effect.

   In the plant species *Physcomitrella patens*, the PcG protein FIE is specifically expressed in the unfertilized egg cell. Soon after fertilisation the FIE gene is inactivated in the young embryo leading to commencement of gene expression (Mosquna *et al.*; 2009).

2. **Post-transcriptional gene silencing** is the result of mRNA of a particular gene being destroyed. The destruction of the mRNA prevents translation to form an active gene product (in most cases, a protein). A **common mechanism of post-transcriptional gene silencing is RNAi**. – (Please refer to your notes on Eukaryotic Gene Regulation for details).

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 10:Multigene families

Introduction to multigene familes and their types

The multigene families are groups of genes of identical or similar sequence. It is a common feature of many eukaryotic genomes. Some of the best known examples of multigene families include those that encode the rRNA, actins, globins, immunoglobulins, tubulins, interferons, histones etc.

Their presence has been highlighted by the techniques of DNA sequencing and genomic studies.

There are two types of criteria to classify the multigene familes.

1. The members of the multigene familes
2. The organization of multigene familes

Based on the members genes, there are two types of multigene familes.

A. Simple or classical multigene familes: In this, all members genes are identical. A well studied example is the rRNA multigene familes. All rRNA genes are present in multiple copies in all eukaryotes. Humans have 280 copies of a repeat unit containing the 28S, 5.8S and 18S rRNA genes and about 2000 gene copies for the 5S rRNA.

B. Complex multigene familes: Such families are made of similar but not identical genes. For example the globin multigene family in humans, tubulin multigene family in plants, *HOX* multigene family in most animal groups etc. In humans, the α-type globins are coded by a small multigene family on chromosome 16 and the ß-type globins by a second family on chromosome 11. Here, the genes in each family are *similar among one another, but not identical*. Hence the resulting proteins have distinctive biochemical properties.

Based on arrangement of multigene families, there are two types.

A. In some multigene families, the individual members are **clustered**, that is they are present in one physical group. For example, the globin genes or the 5S rRNA genes

B. In certain multigene familes, the genes are **dispersed** around the genome, for example the five human genes for *aldolase* are located on chromosomes 3, 9, 10, 16 and 17. The important point is that, even though dispersed, the members of the multigene family have sequence similarities that point to a common evolutionary origin.

Multigene families in plants

It has been proposed that most plant nuclear genes occur in multigene families, which in most cases are the result of gene duplication. However, only a limited number of multigene families have been studied in angiosperms. Well characterized examples include:

1. Nuclear ribosomal genes in *Quercus.*
2. Expansin gene family in most of the plants. Expansins are a group of extracellular proteins that directly modify the mechanical properties of plant cell walls, leading to turgor-driven cell extension. Within the completely sequenced *Arabidopsis* genome, 38 expansin sequences were identified that fall into three discrete subfamilies, α, ß and γ.
3. The two Tubulin gene families in *Arabidopsis*, α and ß tubulins.
4. *adh, leafy*, and calmodulin genes in *Arabidopsis*

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

5. *Vrn* gene family in wheat whose products control vernalisation response
6. Dehydrins gene family with Eleven members in *Hordeum* and six members in *Zea mays*. Dehydrins are a species of protein produced by plants during seed dormancy, drought or low temperature stress.
7. The biggest family consists of 77 genes coding for varieties of ABC transporters—a class of membrane transport proteins found in all three domains of the living world.

Evolutionary distribution of multigene families

Multigene families are known in all groups of organisms, namely Archaea, Prokarya and Eukarya. Several gene families are conserved in different taxa. This provides evidence of evolutionary relatedness.

There was a recent comparison of 2264 gene families in the genomes of 18 bacteria, 6 archaeans and 1 eukaryote (yeast). It was found that 76 are truly ubiquitous (that is, represented in all the genomes analyzed). They are mostly related to functions like:

1. Translation, ribosomal structure and biogenesis
2. Transcription
3. Replication, repair, recombination
4. Cell division and chromosome partitioning
5. Molecule chaperones
6. Outer membrane, cell-wall biogenesis
7. Secretion
8. Inorganic ion transport
9. Signal transduction
10. Energy production and conversion

Origin of multigene families

New genes are generated from preexisting genes.

DNA duplications can generate gene pairs. If both copies are maintained in subsequent generations then a multigene family will exist. Multigene families are believed to have arisen by duplication and variation of a single ancestral gene. In this sense, no gene is ever entirely new. Innovation can occur in four principal ways.

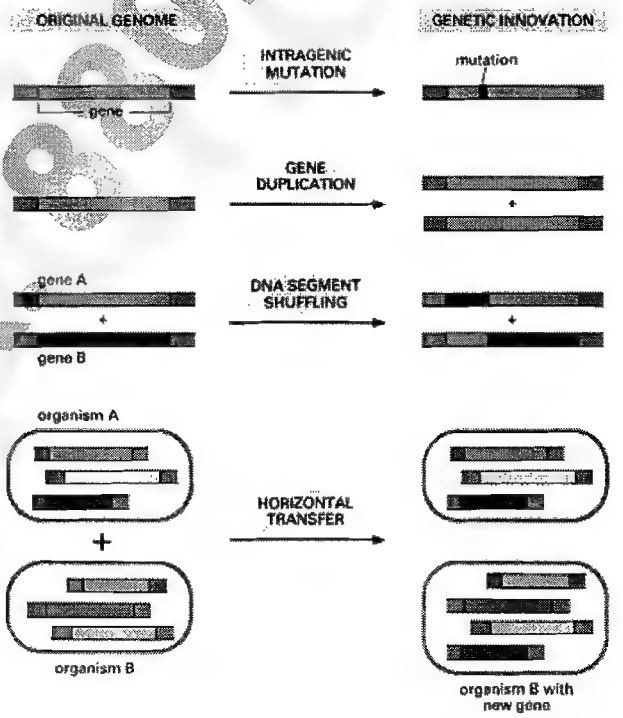


Figure 1: Four Main ways of Genetic Innovation

1. ***Gene duplication***: An existing gene can be duplicated so as to create a pair of closely related genes within a single cell.
2. ***Intragenic mutation***: An existing gene can be modified by mutations in its DNA sequence.
3. ***Gene segment shuffling***: Two or more existing genes can be broken and rejoined to make a hybrid gene consisting of DNA segments that originally belonged to separate genes.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

4. ***Horizontal (intercellular) transfer***: A piece of DNA can be transferred from the genome of one cell to that of another—even to that of another species.

Chromosomal rearrangements disperse the multigene families throughout the genome. Dispersed members of the multigene family can still be recognized by sequence comparison. The evolutionary process described above gives rise to various homology relationships. They are shown in figure 2.

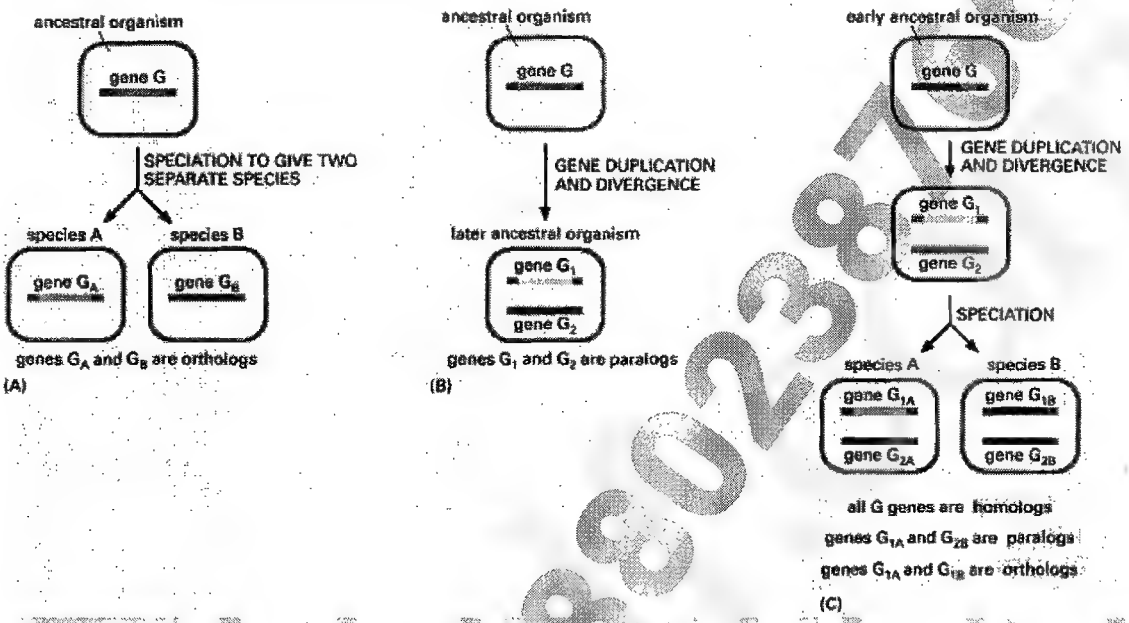


Figure 10: Various types of homologous relationship among genes

Significance of multigene families

1. The significance of recognizing multigene families is that the members may have identical or related functions.
2. It provides insight genes that were derived from a common ancestral gene.
3. It is presumed that many genes are present in multiple copies because there is a heavy demand for rRNA synthesis during cell division, when almost a million of new ribosomes must be assembled.
4. Gene divergence in a complex gene family provides an opportunity for better functional specialization of related genes. The multigene families have been instrumental in complex genomic evolution. One of the best examples is *HOX* multigene family evolution driving the evolution of segmented animals.
5. A comparison of multigene families across organisms gives us reliable information on evolutionary relatedness of organisms and the process of genomic evolution.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 11: Role of RNA in origin and evolution of life

Origin of life

The Earth formed about 4.57 billion years ago, while the earliest fossils of cellular forms available today correspond to a period about 3.5 billion years ago. So **the current view is that the first living form, i.e. the earliest cell, arose from the non-living materials of the early earth about 3.8 – 3.6 bya (billion years ago)**.

Origin of life means the earliest advent of *a self-sustained, self reproducing chemical system capable of undergoing Darwinian evolution* on the earth. Currently, the widely accepted estimates suggest that such a system emerged for the first time on the earth *about 3.8 billion years ago* by ***biochemical abiogenesis***.

There are so many fundamental similarities among all the living forms (in cellular-molecular organization, biochemical pathways, genes & gene expression, molecular genetic pathways *etc.*) that the biologists today conclude that *origin of life occurred only once on the earth and all the living forms today or in the past have ultimately descended from a single common universal ancestor*.

Biochemical Abiogenesis Model

Biochemical abiogenesis model explains how biological life arose from inorganic matter through natural processes of increasing chemical complexity. Abiogenesis occurred between 3.9 and 3.5 billion years ago.

In 1924, Alexander Ivanovich Oparin proposed that simple one-celled life forms might have come from the simple organic molecules present in the early earth's atmosphere. From geological evidences, he proposed theoretical conditions for the early earth. The early atmosphere was quite different from the modern atmosphere and probably resembled atmospheres of other planets.

In the early atmosphere the following chemicals abounded but there was no free oxygen gas.

1. Hydrogen
2. Ammonia
3. Methane
4. Carbon Dioxide
5. Water
6. Nitrogen

Oparin proposed that in the extreme primitive conditions it was possible that these simple molecules combined with one another to give rise to the bio-organic macromolecules. Later, these bio-organic macromolecules might have combined in a self-replicating system and gave rise to the earliest cells.

In 1928, English biochemist J.B.S. Haldane proposed that life might have arisen on the earth when Oparin's early atmosphere was subjected to energy inputs in the forms of Heat from the cooling earth and Ultraviolet radiation. Haldane also said that lightening might have provided the sudden but large inputs of additional energy for attaining biochemical complexity.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

In 1953, Stanley Miller, a graduate student in biochemistry, built the apparatus shown here. He filled it with:

1. Methane (CH_4);
2. Ammonia (NH_3);
3. Water (H_2O);
4. Hydrogen (H_2); but no oxygen.

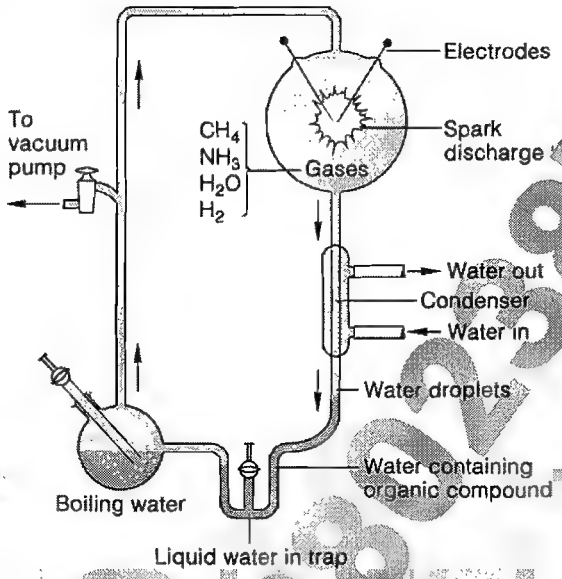


Figure 1: Miller's Experiment

He hypothesized that this mixture resembled the atmosphere of the early earth. The mixture was kept circulating by continuously boiling and then condensing the water. The gases passed through a chamber containing two electrodes with a spark passing between them. At the end of a week, Miller used paper chromatography to show that the flask now contained several amino acids, purines, pyrimidines as well as some other organic molecules.

This experiment was the validation of the earlier two theories.

Role of RNA

When the structure of DNA was described by Watson and Crick in 1953, it was proposed that DNA was the first dominant biomolecule during the early steps towards the origin of life. The basis of this argument was that DNA has a self-complementary double stranded structure which allowed it to replicate. Before this proposal, the proteins were believed to be the earliest dominant biomolecules due to their versatile functional nature.

However, both the molecules speculated to be the early dominant molecules had certain limitations. First of all, DNA – despite a self-complementary double stranded structure – cannot replicate without the catalytic assistance of a number of proteins. Similarly, the limitation with proteins is that they are constructed by linking amino acids under the genetic instruction of DNA.

Thus, it was obvious that neither DNA nor Proteins could be the earliest class of dominant biomolecules.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

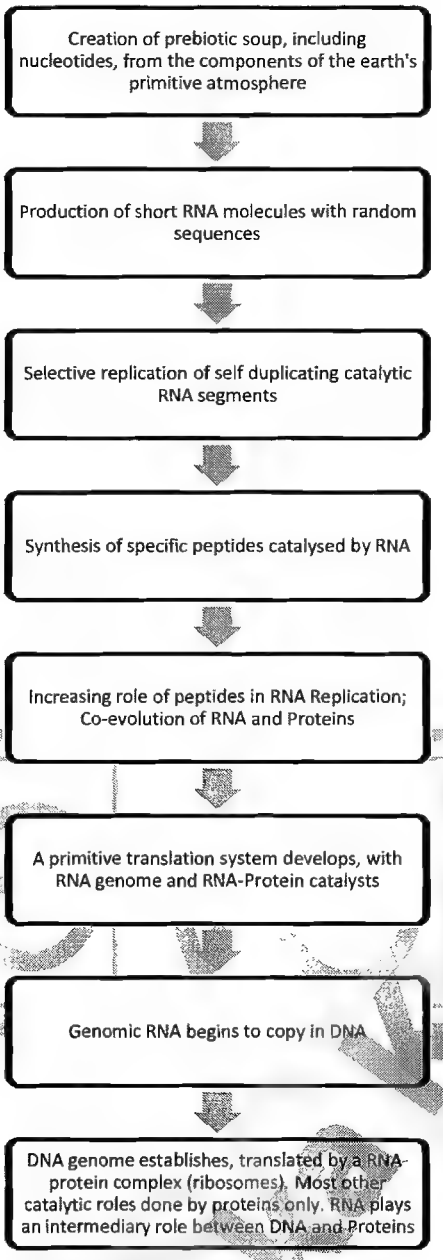


Figure 2: Role of RNA in origin of life

In mid-1980, Sidney Altman and Thomas Cech separately published their discovery of self-splicing (catalytic) RNA. Thus, finally a class of biomolecule was identified which could store genetic information as well as could carry out catalytic functions. In 1986, Walter Gilbert coined the term **RNA World** to describe a time during which RNA was the primary information and catalytic molecule.

The RNA world hypothesis suggests that short RNA molecules could have spontaneously formed that would then catalyze their own continuing replication.

A proposed alternative to the RNA World is called PNA World, dominated by peptide nucleic acids. PNA is more stable than RNA and appears to be more readily synthesised in prebiotic conditions, where especially the synthesis of ribose and adding phosphate groups are problematic.

The transition between the PNA world and the RNA world probably occurred as the conditions on the earth began to be more benign. The plausibility of this scheme is supported by laboratory experiments showing that PNA can act as a template for the synthesis of complementary RNA molecules.

In any self replicating system, Darwinian natural selection is bound to operate.

In a mature RNA world, different forms of RNA competed with each other for free nucleotides, and were subjected to natural selection. The most efficient molecules of RNA, the ones able to catalyze their own reproduction, survived and finally gave rise to the modern RNA. It is possible that competition between RNA favored the emergence of cooperation between RNA molecules opening the way to more complex supra molecular assemblies.

DNA, Proteins and other biomolecules were recruited later into life.

The possible sequence of events is shown below in Figure 2.

Support for the RNA world model

An *RNA world* existed on Earth before modern cells arose. According to this hypothesis, RNA stored both genetic information and catalyzed the chemical reactions in primitive cells. Only later in evolutionary time, DNA took over as the genetic material and proteins become the major catalyst and structural component of cells.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

The **RNA world hypothesis** proposes that RNA was, before the emergence of the first cell, the dominant, and probably the only, form of life on Earth.

This hypothesis is supported by RNA's multiple abilities.

- To participate in the storage, transmission, and duplication of genetic information (similarly to DNA)
- To act as a ribozyme (similar to an enzyme), catalyzing certain reactions.

Thus, from the point of view of reproduction, molecules exist for two basic purposes: self-replication and catalysis assisting self-replication. DNA is capable of self-replication, but only assisted by proteins. Proteins are excellent catalysts, but fail to catalyze processes complex enough to recreate themselves, individually. RNA is capable of both catalysis and self-replication, which is the biggest support for the RNA World Hypothesis.

In 2001, the RNA world hypothesis was given a major boost with the deciphering of the 3-dimensional structure of the ribosome, which revealed that the key catalytic sites of the ribosome were composed of RNA, with proteins playing only a structural role in holding the ribosomal RNA together. Specifically, the formation of the peptide bond, the reaction that binds amino acids together into proteins, is catalyzed solely by RNA. This finding suggests that RNA molecules were responsible for (or at the very least capable of) generating the first proteins.

Other abilities of RNA also support the model. They include:

- To undergo allosteric conformational changes, either in response to small molecules or to other RNAs
- To fold up in many various ways and perform various tasks. Even today RNA plays a number of intermediary role between DNA and Proteins:
    1. Primer formation for DNA replication
    2. Telomere replication by Telomerase activity
    3. Transcription
    4. Splicing and Alternate splicing
    5. mRNA editing by Guide RNA system
    6. Heterochromatization of X chromosome [XIST]
    7. Translation [Peptidyl Transferase Centre]
    8. Riboswitches to regulate Transcription and Translation
    9. RNAi
    10. 7SL RNA in Protein Targeting to the RER lumen for core glycosylation
- Many of the cofactors that play so many roles in life are based on ribose; for example:
    1. ATP
    2. NAD

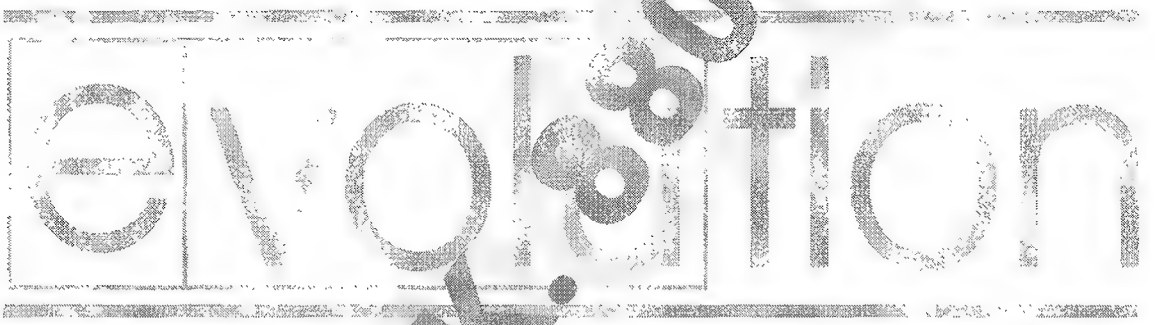
Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

3. FAD
4. coenzyme A
5. cyclic AMP
6. GTP

Conclusion

From our knowledge of present-day organisms and the molecules they contain, it seems likely that the development of the directly autocatalytic mechanisms fundamental to living systems began with the evolution of families of molecules that could catalyze their own replication. With time, a family of cooperating RNA catalysts probably developed the ability to direct synthesis of polypeptides. DNA is likely to have been a late addition: as the accumulation of additional protein catalysts allowed more efficient and complex cells to evolve, the DNA double helix replaced RNA as a more stable molecule for storing the increased amounts of genetic information required by such cells.



Priyanka K. 8802338567

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Chapter 12: Mutations

What is a mutation?

A mutation is a heritable change in the genetic information within a cell that has arisen as a random and single event in the genome. Classically, a mutation was defined as a ***sudden and heritable change in a phenotypic trait***. However, in modern definition there is an emphasis on ***heritable change in genetic material*** **because** it encompasses mutations of all kinds – including those which have no apparent phenotypic result (e.g. Silent mutation).

A mutation is a random process. It means that the individual or the cell suffering a mutation has no control over the location, incidence or magnitude of the mutation it suffers. It is also a non adaptive process. It means that organisms can not decide to mutate their genetic material in order to adapt to a new condition.

Classification

Mutations have been classified on various grounds – of which some important ones are described below.

1. By extent

Based on extent, there are two basic types of mutations.

1. Mutations that modify only a single gene are called **gene mutations**. The gene mutations are also called **point mutations** or **micro mutations** and they involve only one or few nucleotides. These mutations can not be observed by morphological examination of the chromosome. To study gene mutation the geneticists normally investigate their biochemical or phenotypic outcome.
2. Mutations that modify blocks of genes on a chromosome are called **chromosomal mutations** (or **Gross Mutations** or **Macro mutations**). These mutations are readily studied by morphological examination of the chromosome. Gross mutations are either **structural aberrations of the chromosome** (for example gross insertion, gross deletions, translocations, duplications etc.) or **ploidy mutations**. In modern genetics, the gross chromosomal mutations are studied under a separate branch of structural and numerical aberrations of chromosome.

Point and Gross mutations can further be classified, as described below.

- Point mutations affecting one or a few nucleotides, include –
    - **Base substitution** in which a nitrogenous base is replaced by some other base. Base substitutions are of two types.
        A. **Transitions** in which a purine is replaced by a purine and a pyrimidine is replaced by a pyrimidine.
        B. **Transversions** where a purine is replaced by a pyrimidine and vice versa.
    - **Insertions** add one or more extra nucleotides into the DNA. They are usually caused by transposable elements, or errors during replication of repeating elements (e.g. AT repeats). These mutations can alter the reading frame of the gene.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

- **Deletions** remove one or more nucleotides from the DNA. Like insertions, these mutations can also alter the reading frame of the gene. They are irreversible.

Note: The mutations which can alter the reading frame of the gene are called Frame Shift Mutations.

- Large-scale mutations in chromosomal structure, include –
  - **Amplifications** (or gene duplications) leading to multiple copies of chromosomal regions, increasing the dosage of the genes located within them.
  - **Triplet Expansions** Some triplet repeats of substantial length are prone to elongation in larger steps that the addition of single repeats, leading to massive enlargement. **Triplet expansion** is the cause of several human genetic diseases including Huntington Disease. Strictly speaking, triplet expansions are types of Amplification.
  - **Deletions** of large chromosomal regions, leading to loss of the genes within those regions.
  - Mutations whose effect is to juxtapose previously separate pieces of DNA, by bringing together separate genes to form functionally distinct fusion genes. These include:
    - **Chromosomal translocations**: attaching DNA from separate chromosomes.
    - **Interstitial deletions**: removing regions of DNA from a single chromosome, thereby apposing previously distant genes.
    - **Chromosomal inversions**: switching the orientation of a segment of a chromosome, thereby apposing its ends to previously distant genes.
  - **Loss of heterozygosity**: loss of one allele, either by a deletion or recombination event, in organisms which previously had two.

2. By the tissue affected

In multicellular organisms, we can distinguish between two broad categories of mutations: somatic mutations and germ-line mutations.

- **Somatic mutations** arise in somatic tissues, which do not produce gametes. When a somatic cell is with a mutation, the mutation is passed on only to the daughter cells leading to a population of genetically identical mutant cells. The earlier in development that a somatic mutation occurs, the larger is the spread of the mutant cells.
- **Germ-line mutations** arise in cells that ultimately produce gametes. A germ-line mutation – if introduced early during the germ line differentiation – can be passed to future generations, producing organisms that carry the mutation in all their somatic and germ-line cells.

3. By effect on function

Gene mutations have varying effects on health, depending on where they occur and whether they alter the function of essential proteins.

Point mutations that occur wihtin the protein coding region of a gene may be classified into four kinds, depending upon what the erroneous codon codes for:

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

- **silent mutations**: codes for the same amino acid
- **neutral mutations**: changes the amino acid sequence of a protein without altering its ability to function
- **mis-sense mutations**: codes for a different amino acid
- **nonsense mutations**: codes for a stop, which can truncate the protein
- **null mutations**: where the protein is not synthesized at all or the protein is completely non functional

Other typifications use some more terms:

- **Loss-of-function mutations** are the result of gene product having less or no function. When the allele has a complete loss of function (null allele) it is often called an **amorphic mutation**. Phenotypes associated with such mutations are most often recessive. Exceptions are when the organism is haploid, or when the reduced dosage of normal gene product is not enough for normal phenotype (this is called **haploinsufficiency**).
- **Gain-of-function mutations** change the gene product such that it gains a new and abnormal function. These mutations usually have dominant phenotypes.
- **Dominant negative mutations** (also called **antimorphic mutations**) have an altered gene product that acts antagonistically to the wild-type allele. These mutations usually result in an altered molecular function (often inactive) and are characterised by a dominant or semi-dominant phenotype.
- **Lethal mutations** are mutations that lead to a phenotype incapable of effective reproduction.

4. By effect on the pattern of gene expression

- **Hypomorphic mutations** are mutations that cause reduced function of the gene product, or a negative change in expression of the gene.
- **Hypermorphic mutations** are the opposite of hypomorphic mutations; they cause increased activity or expression of the gene product.
- **Neomorphic mutations** cause a novel molecular function or expression of the gene product.
- **Conditional mutation** is a mutation that has wild-type phenotype under certain environmental conditions and a mutant phenotype under certain selective conditions. Conditional mutations may also be lethal.

5. By aspect of phenotype affected

- **Morphological mutations** usually affect the outward appearance of an individual. Mutations can change the height of a plant or change it from smooth to rough seeds.
- **Biochemical mutations** result in lesions stopping the enzymatic pathway. Often, morphological mutants are the direct result of a mutation due to the enzymatic pathway.

6. By other genes affected

- **Polar mutations** when the expression of even neighbouring genes is affected

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

- **Non Polar mutations** when the expression of neighbouring genes is not affected

7. By reversibility

- **Reversible**, which can reverse itself by simple molecular process
- **Irreversible**, which can not reverse itself by simple molecular processes.

8. By spontaneity

Two classes of mutations on this basis are spontaneous mutations (molecular decay) and induced mutations caused by mutagens.

- **Spontaneous mutations** on the molecular level include the errors in DNA replicative process (error in proofreading), slipped strand, or metabolic accidents etc. The DNA has so-called **hotspots**, where mutations occur up to 100 times more frequently than the normal mutation rate. A hotspot can be at an unusual base, e.g., 5-methylcytosine. Contrastingly, there are mutational **cold spots** as well.
- **Induced mutations** on the molecular level can be caused by:
  - Chemicals
    - Nitrosoguanidine (NTG)
    - Base analogs (e.g. BrdU)
    - Simple chemicals (e.g. acids)
    - Alkylating agents (e.g. *N*-ethyl-*N*-nitrosourea (ENU)) These agents can mutate both replicating and non-replicating DNA. In contrast, a base analog can only mutate the DNA When the analog is incorporated in replicating the DNA. Each of these classes of chemical mutagens has certain effects that then lead to transitions, transversions, or deletions.
    - Methylating agents (e.g. ethane methyl sulfonate (EMS))
    - Polycyclic hydrocarbons (e.g. benzpyrenes found in internal combustion engine exhaust)
    - DNA intercalating agents (e.g. ethidium bromide)
    - DNA crosslinker (e.g. platinum)
    - Oxidative damage caused by oxygen radicals
  - Radiations such as Ultraviolet radiation and Ionizing radiation

9. By direction

A mutation that alters the wild-type phenotype is called a **forward mutation**, where as a **reverse mutation** (a reversion) changes a mutant phenotype back into the wild type.

10. By effect on other mutations

- A **suppressor mutation** is a genetic change that hides or suppresses the effect of another mutation. This type of mutation is distinct from a reverse mutation, in which the mutated site changes back into the original wild-type sequence. A suppressor mutation occurs at a site

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

that is distinct from the site of the original mutation. There are two types of suppressor mutations. A. **Intragenic suppressor:** Suppresses the effect of an earlier mutation within the same gene mutation B. **Intergenic suppressor:** Suppresses the effect of an earlier mutation in another gene

- A **limited mutation** has no effect on other mutations.
- An **enhancer mutation** enhances the effect of some other mutation.

11. By dominance relation

- **Majority of the mutations are recessive** with respect to the wild type allele.
- However **some rare mutations are dominant**. Notable examples of dominant mutations include: *notch* wing mutation in *Drosophila*, Vitamin D resistant Rickets in Humans, Hereditary enamel hypoplasia in humans, and Achondroplastic dwarfness in humans.

12. By effect on survival

- **Lethal mutations** kill the individual carrying them.
- **Sub-lethal mutations** do not kill the individual but lead to some serious biochemical or physiological impairment.
- **Conditional lethal mutations** kill the individual carrying them only under some well defined conditions.

TABLE 1: Summarized characteristics of some important types of mutations

Type of mutation	Definition
Base substitution	Changes the base of a single DNA nucleotide
Transition	Base substitution in which a purine replaces a purine or a pyrimidine replaces a pyrimidine
Transversion	Base substitution in which a purine replaces a pyrimidine or a pyrimidine replaces a purine
Insertion	Addition of one or more nucleotides
Deletion	Deletion of one or more nucleotides
Frameshift mutation	Insertion or deletion that alters the reading frame of a gene
In-frame mutation	Insertion or deletion of a multiple of three nucleotides or insertion that or insertion does not alter the reading frame
Expanding trinucleotide	Repeated sequence of three nucleotides (trinucleotide) in which the repeats number of copies of the trinucleotide increases
Forward mutation	Changes the wild-type phenotype to a mutant phenotype
Reverse mutation	Changes a mutant phenotype back to the wild-type phenotype
Missense mutation	Changes a sense codon into a different sense codon, resulting in the incorporation of a different amino acid in the protein
Nonsense mutation	Changes a sense codon into a nonsense codon, causing premature termination of translation

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Silent mutation	Changes a sense codon into a synonymous codon, leaving unchanged the amino acid sequence of the protein
Neutral mutation	Changes the amino acid sequence of a protein without altering its ability to function
Loss-of-function mutation	Causes a complete or partial loss of function
Gain-of-function mutation	Causes the appearance of a new trait or function or causes the appearance of a trait in inappropriate tissue or at an inappropriate time
Lethal mutation	Causes premature death
Suppressor mutation	Suppresses the effect of an earlier mutation at a different site
Intragenic suppressor	Suppresses the effect of an earlier mutation within the same gene mutation
Intergenic suppressor	Suppresses the effect of an earlier mutation in another gene.

Mutation Rates & Mutation Frequencies

The frequency with which a gene changes from the wild type to a mutant is referred to as the **mutation rate** and is generally expressed as the number of mutations per biological unit, which may be mutations per cell division, per gamete, or per round of replication. For example, the mutation rate for achondroplasia (a type of hereditary dwarfism) is about four mutations per 100,000 gametes, usually expressed more simply as 4×10^{-5}.

Mutation rates vary widely across species.

In contrast, **mutation frequency** is defined as the incidence of a specific type of mutation within a group of individual organisms. For achondroplasia, the mutation frequency in the United States is about 2×10^{-4}, which means that about 1 of every 20,000 persons in the U.S. population carries this mutation. Thus mutation frequency is a measure of incidence frequency of a particular mutation within a defined population.

Mechanisms of gene mutation

Mutations result from both internal and external factors. Those that are a result of natural changes in DNA structure are termed **spontaneous mutations**, whereas those that result from changes caused by environmental chemicals or radiation are **induced mutations**.

Spontaneous Mutagenesis

1. Spontaneous Replication Errors

Replication is highly accurate with an error tolerance of not more than 10^{-9} - that is one in a billion. However, spontaneous replication errors occasionally occur and lead to mutation. There are four major types of replicational mutagenesis.

1. The primary cause of spontaneous replication errors is **tautomeric shifts**, in which the positions of protons in the DNA bases change (Fig. 1). Purine and pyrimidine bases exist in different chemical forms called tautomers. The two tautomeric forms of each base are in dynamic equilibrium although one form is more common than the other. The standard Watson and Crick base pairings—adenine with thymine, and cytosine with guanine—are

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

between the common forms of the bases, but, if the bases are in their rare tautomeric forms, other base pairings are possible.

When a mismatched base has been incorporated into a newly synthesized nucleotide chain, an **incorporated error** is said to have occurred. This incorporated error leads to mutation, in the next round of DNA Replication. It is illustrated in the example below.

Suppose that, in replication thymine (which normally pairs with adenine) mispairs with guanine through rare tautomeric shift. In the next round of replication the two mismatched bases separate, and each serves as template for the synthesis of a new nucleotide strand. This time, if thymine is back to its normal tautomeric form, it pairs with adenine, producing another copy of the original DNA sequence. On the other strand, however, the incorrectly incorporated guanine serves as the template and pairs with cytosine, producing a new DNA molecule that has an error—a C·G pair in place of the original T·A pair (a T·A→C·G base substitution). **The original incorporated error leads to a replication error, which creates a permanent mutation**, because all the base pairings are correct and there is no mechanism for repair systems to detect the error.

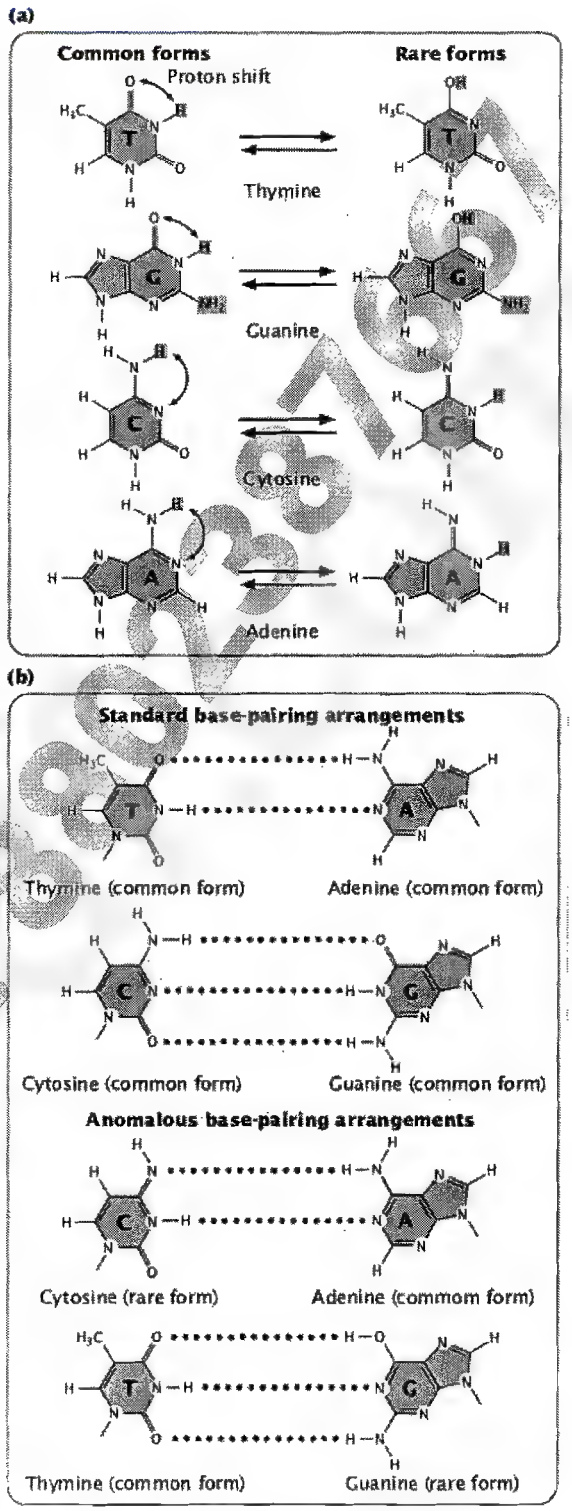


Figure 1: Tautomeric shifts

2. Mispairing can also occur through **wobble**, in which normal, protonated, and other forms of the bases are able to pair because of flexibility in the DNA helical structure. These structures have been detected in DNA molecules and are now thought to be responsible for many of the mispairings in replication. Wobble leads to an incorporated error. As described earlier, this incorporated error leads to mutation, in the next round of DNA Replication.

3. Mutations due to **small insertions and deletions** also may arise spontaneously in replication and crossing over. Strand slippage may occur when one nucleotide strand forms a

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

small loop. If the looped-out nucleotides are on the newly synthesized strand, an insertion results. At the next round of replication the insertion will be incorporated into both strands of the DNA molecule. If the looped-out nucleotides are on the template strand, then the newly replicated strand has a deletion, and this deletion will be perpetuated in subsequent rounds of replication.

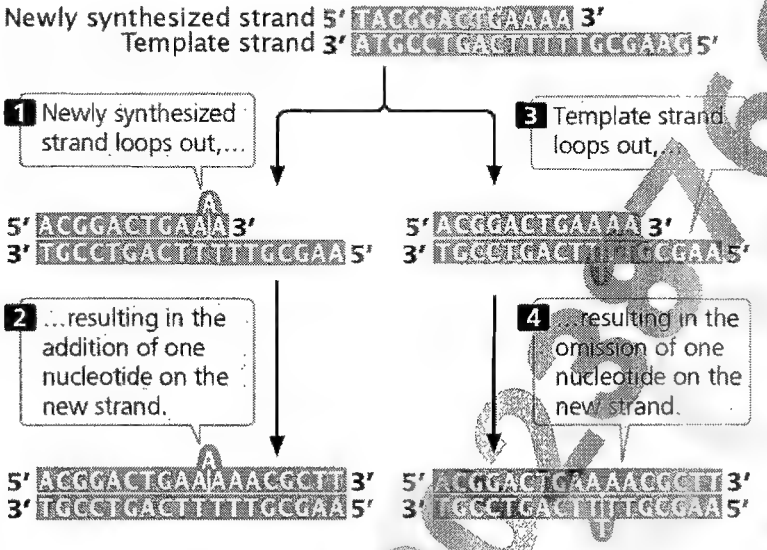


Figure 2: Insertions and deletions arising from strand slippage

4. Misaligned pairing may cause **unequal crossing over**, which results in one DNA molecule with an insertion and the other with a deletion.

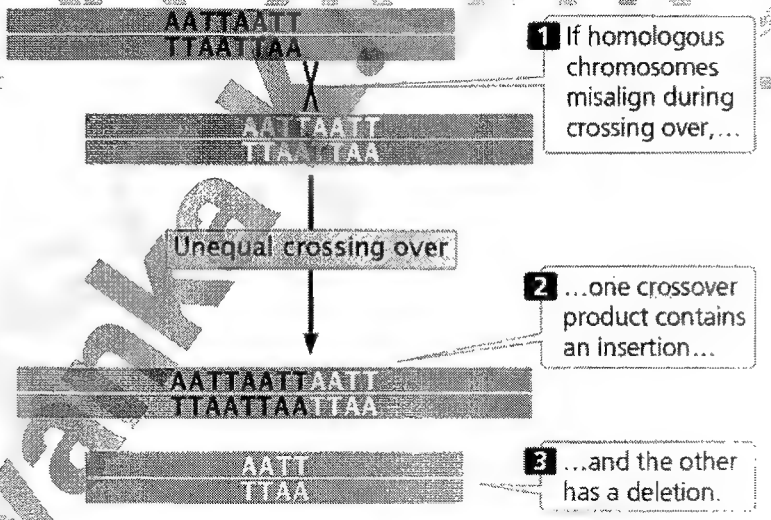


Figure 3: Unequal Crossing Over

2. Spontaneous chemical changes

In addition to spontaneous mutations that arise in replication mutations also result from spontaneous chemical changes in DNA.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

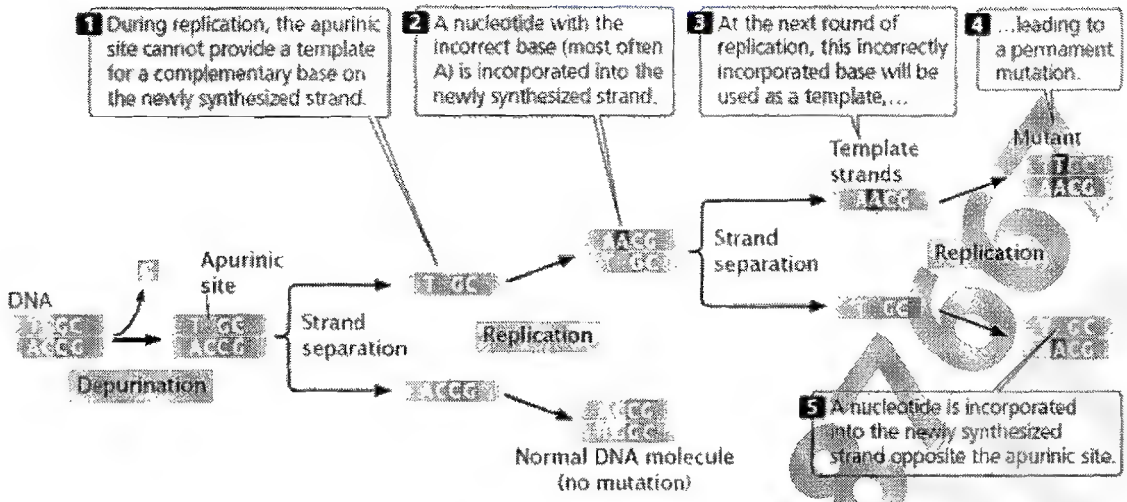


Figure 4: Depurination

1. **Depurination:** Depurination is the loss of a purine base from a nucleotide. Depurination results when the covalent bond connecting the purine to the carbon atom of the deoxyribose sugar breaks, producing an apurinic site—a nucleotide that lacks its purine base. An apurinic site cannot act as a template for a complementary base in replication. In the absence of base-pairing constraints, an incorrect nucleotide (most often adenine) is incorporated into the newly synthesized DNA strand opposite the apurinic site, frequently leading to an incorporated error. The incorporated error is then transformed into a replication error at the next round of replication (Fig.4). Depurination is a common cause of spontaneous mutation; a mammalian cell in culture loses approximately 10,000 purines every day.

2. **Deamination:** Another spontaneously occurring chemical change that takes place in DNA is deamination, the loss of an amino group (NH_2) from a base. Deamination may occur spontaneously or be induced by mutagenic chemicals like Nitrous Acid. Deamination may alter the pairing properties of a base: **the deamination of cytosine, for example, produces uracil**, which pairs with adenine during replication. After another round of replication the adenine will pair with thymine, creating a T·A pair in place of the original C·G pair (C·G→U·A→T·A); this chemical change is a transition mutation. This type of mutation is usually repaired by enzymes that remove uracil whenever it is found in DNA. **Deaminationation of Adenine forms Hypoxanthine**, which behaves as Guanine. **Deaminated guanine is xanthine**, which pairs with cytosine just as guanine does; however, xanthine may also pair with thymine, leading to a C·G→T·A transition.

Induced Mutagenesis

1. Chemically Induced Mutations

Although many mutations arise spontaneously, a number of environmental agents are capable of damaging DNA, including certain chemicals and radiation. **Any environmental agent that significantly increases the rate of mutation above the spontaneous rate is called a mutagen.**

Important modes of chemically induced mutagenesis are listed and described below.

1. **Mutagenesis by Base analogs:** Base analogs are chemicals with structures similar to that of any of the four standard bases of DNA. DNA polymerases cannot distinguish these analogs

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

from the standard bases; so, if base analogs are present during replication they may be incorporated into newly synthesized DNA molecules. For example, **5–bromouracil (5BU) is an analog of thymine**. Normally, 5–bromouracil pairs with adenine just as thymine does, but it occasionally due to tautomeric shift mispairs with guanine, leading to a transition (T·A→BU·A→BU·G→C·G). Through mispairing, 5–bromouracil may also be incorporated into a newly synthesized DNA strand opposite guanine (Fig. 5).

Another mutagenic chemical is **2–aminopurine (2AP), which is a base analog of adenine**. Normally, 2–aminopurine base pairs with thymine, but it may mispair with cytosine, causing a transition mutation (T·A→T·2AP→C·2AP→C·G). Alternatively, 2–aminopurine may be incorporated through mispairing into the newly synthesized DNA opposite cytosine and then later pair with thymine, leading to a C·G→C·2AP→T·2AP→T·A transition.

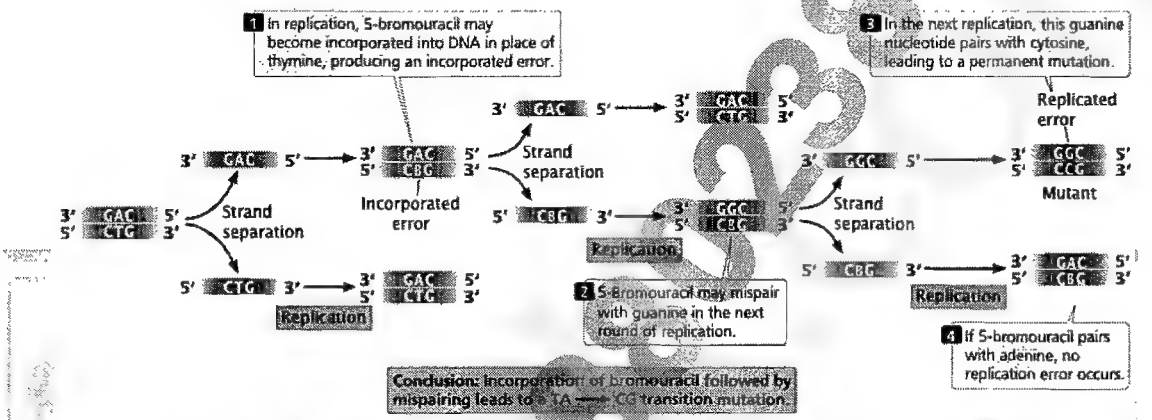


Figure 5: Mutation due to base analog

2. **Alkylating agents** Alkylating agents are chemicals that donate alkyl groups, such as methyl (CH_3) and ethyl (CH_3–CH_2) groups, to nucleotide bases. For example, **ethylmethanesulfonate (EMS) adds an ethyl group to guanine, producing 6-ethylguanine, which pairs with thymine**. Thus, EMS produces C·G→T·A transitions. **EMS is also capable of adding an ethyl group to thymine, producing 4-ethylthymine, which then pairs with guanine**, leading to a T·A→C·G transition. Because EMS produces both C·G→T·A and T·A→C·G transitions, mutations produced by EMS can be reversed by additional treatment with EMS. Mustard gas is another alkylating agent.

   Many a times alkylating agents also contribute to the creation of AP sites, whose consequences are already described.

3. **Deamination:** In addition to its spontaneous occurrence, deamination can be induced by some chemicals. For instance, nitrous acid deaminates cytosine, creating uracil, which in the next round of replication pairs with adenine, producing a C·G→T·A transition mutation. Nitrous acid changes adenine into hypoxanthine, which pairs with cytosine, leading to a T·A→C·G transition. Nitrous acid also deaminates guanine, producing xanthine, which pairs with cytosine just as guanine does; however, xanthine may also pair with thymine, leading to a C·G→T·A transition. Nitrous acid produces exclusively transition mutations and, because both C·G→T·A and T·A→C·G transitions are produced, these mutations can be reversed with nitrous acid.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

4. **Hydroxylamine:** Hydroxylamine is a very specific base-modifying mutagen that **adds a hydroxyl group to cytosine, converting it into hydroxylaminocytosine**. This conversion increases the frequency of a rare tautomer that pairs with adenine instead of guanine and leads to C·G→T·A transitions. Because hydroxylamine acts only on cytosine, it will not generate T·A→C·G transitions; thus, hydroxylamine will not reverse the mutations that it produces.

5. **Oxidative reactions**: Reactive forms of oxygen (including superoxide radicals, hydrogen peroxide, and hydroxyl radicals) are produced in the course of normal aerobic metabolism as well as by radiation, ozone, peroxides, and certain drugs. These reactive forms of oxygen damage DNA and induce mutations by bringing about chemical changes to DNA. For example, **oxidation converts guanine into 8-oxy-7,8-dihydrodeoxyguanine, which frequently mispairs with adenine instead of cytosine**, causing a G·C→T·A transversion mutation.

6. **Intercalating agents**: Proflavin, acridine orange, ethidium bromide, and dioxin are intercalating agents, which are about the same size as a nucleotide. Intercalating agents produce mutations by sandwiching themselves (intercalating) between adjacent bases in DNA, distorting the three-dimensional structure of the helix and causing single-nucleotide insertions and deletions in replication. These insertions and deletions frequently produce frameshift mutations (which change all amino acids downstream of the mutation), and so the mutagenic effects of intercalating agents are often severe. Because intercalating agents generate both additions and deletions, they can reverse the effects of their own mutations.

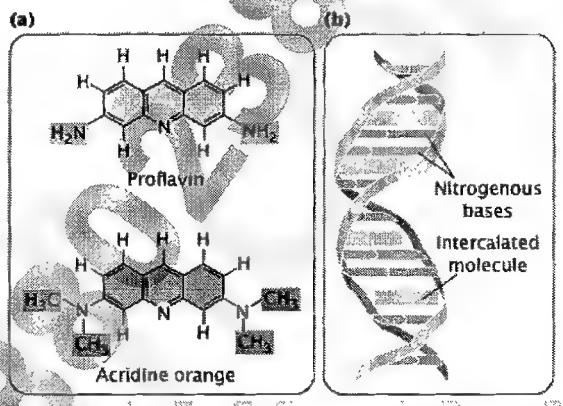


Figure 6: effect of an intercalating agent

2. Radiation Induced Mutations

In 1927, Hermann Muller demonstrated that mutations in fruit flies could be induced by X-rays. The results of subsequent studies showed that X-rays greatly increase mutation rates in all organisms.

It is now known that many kinds of radiation increase mutations. Table 2 shows the effects of several types of radiation on increasing mutation frequencies in *Drosophila.*

Table 2. Relative Efficiencies of Various Types of Radiation in Producing Mutations in *Drosophila*

Type of radiation	Percentage of male X chromosomes bearing recessive lethal mutations after a dose of 1000 roentgens*
Visible light	0.15
X rays (25 Mev)	1.70
β rays, γ rays	2.90
Hard X rays	2.90

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

Soft X rays	2.50
Neutrons	1.90
α rays	0.84
UV ray A	0.76
UV ray B	1.24
UV ray C	1.96
* The roentgen (r) is a unit of radiation energy.	

Radiation is often categorized as *ionizing* or *nonionizing*, depending on whether ions are produced in the tissue through which it passes. X rays and γ (gamma) rays, β rays, α rays and cosmic rays produce ions, and UV radiation does not. The high energies of X rays and γ (gamma) rays, β rays, α rays and cosmic rays are all capable of penetrating tissues and damaging DNA. These forms of radiation, called ionizing radiation, can:

1. dislodge electrons from the atoms that they encounter, changing stable molecules into free radicals and reactive ions, which can show mispairing
2. alter the structures of bases
3. break phosphodiester bonds in DNA.
4. cause double-strand breaks in DNA. Attempts to repair these breaks can produce chromosome mutations .

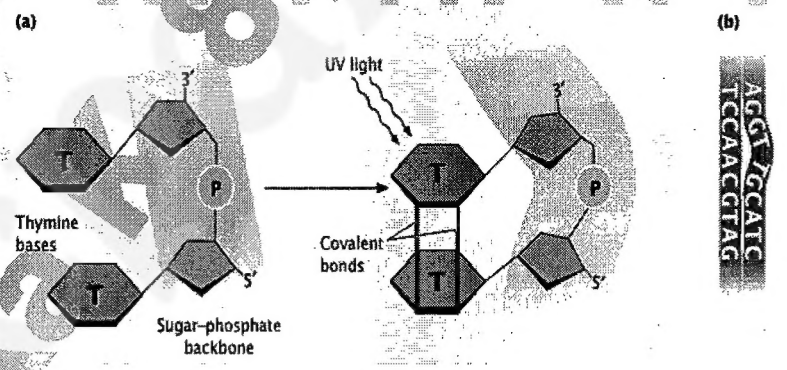


Figure 7: Formation of thymine dimers

The ionizing radiations cause the formation of ionized nitrogenous bases, which have a tendency to mispair (just the way the rare tautomeric forms of the bases have). As a result of base ionization, thus, incorporational errors occur during DNA replication which later on lead to mutations.

On the other hand, Ultraviolet radiation has less energy than that of ionizing radiation. Hence it does not eject electrons and cause ionization but is nevertheless highly mutagenic. The method of UV ray induced mutagenesis is explained below.

Purine and pyrimidine bases readily absorb UV. This results in the formation of chemical bonds between adjacent pyrimidine molecules on the same strand of DNA and in the creation of structures called pyrimidine dimers. Pyrimidine dimers consisting of two thymine bases (called thymine dimers) are most frequent (Fig. 6), but cytosine dimers and thymine–cytosine dimers also can form. These dimers distort the configuration of DNA and often block replication. With a low frequency, this leads to deletional mutations. Very often, when pyrimidine dimers block

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

replication cell division is inhibited and the cell usually dies; for this reason, UV light kills bacteria and is an effective sterilizing agent used in water purification devices such as Aqua-Guard™.

Most pyrimidine dimers are immediately repaired by various mechanisms.

A summary of important mutagens

1. Chemicals causing mutations

Deaminating Agents

Molecules inducing base deamination, including nitrous acid which is nonspecific and sodium bisulfite, which specifically deaminates cytosine & Nitrous acid.

Alkylating Agents

Molecules reacting with nucleophilic centres in nucleic acids and substituting them with alkane groups and their derivatives. Such chemicals are often potent mutagens; examples include ethyl methane sulfonate (EMS) and dimethylnitrosamine. Alkylating agents cause various types of DNA lesion - base modification may change the base pairing potential of the alkylated base causing a misincorporation during replication; alternatively a bulky adduct may block replication. Bifunctional agents (those which can alkylate two nucleophilic centres) may form cross-links. Other examples include: Mustard gas, Nitrogen mustard, MMS, EES and N-methyl-N'-nitro-N'nitrosoguanidine.

Donors of Bulky Groups

Molecules which, when metabolized, form reactive species which add bulky chemical groups to bases. Examples include aflatoxin B1 and benzopyrene. The consequence of such a reaction is a bulging distorted helix which blocks replication.

Base Analogs

Molecules resembling bases which form nucleotides that can be incorporated into a growing nucleotide chain (e.g. bromodeoxyuracil). Unlike the major bases, base analogs may demonstrate aberrant base pairing properties (e.g. they may pair with more than one base because they are stable in more than one tautomeric form). Base analogs generate point mutations, often in a highly specific manner. Other examples include: 2-aminopurine, 5-iodouracial, 2,6 –diamino-purine, and 5-chlorouracil.

Intercalating Agents

Molecules with a planar component which fits in between the bases of DNA (e.g. ethidium bromide, acridine orange, proflavin, ICR 170, ICR 191). Intercalating agents increase the length of the DNA strand by unwinding it. This induces frameshifts, blocks replication and inhibits nucleotide excision repair by sequestering the enzymes into inactive complexes.

Crosslinking Agents

Molecules facilitating the covalent attachment of DNA strands. Such agents include bifunctional alkylating agents, nitrogen and sulfur mustards and platinum derivatives. Planar molecules, Psoralens, also cross-link DNA bases when exposed to UV light.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf

2. Physical mutagenic agents

Ionizing Radiation

Ionizing radiation causes a variety of DNA lesions including damaged bases, damaged sugar rings, nicks and breaks. The effects of ionizing radiation may be direct (caused by ionization of atoms within the DNA molecule) or indirect (caused by the generation of other reactive molecules in the cell, predominantly reactive oxygen species, which then interact with DNA).

UV Radiation

UV-induced lesions are termed photoproducts. The predominant effect of UV irradiation is the generation of photodimers involving adjacent bases. The most common photodimer is the cyclobutyl pyrimidine dimer, occurring between any two adjacent pyrimidines, T = T being most frequent, followed by C= T, T =C and C=C. Another frequently observed photodimer is the (6-4) lesion, and photodimers involving purines are also formed. UV radiation may. also induce damage to single bases, often by hydrating them. A unique spore photoproduct is generated by UV irradiation of *B. subtilis* spores

3. Biological mutagenic agents

Mobile Genetic Agents

DNA sequences which can move around in the genome. Examples include viruses (e.g. bacteriophage Mu, retroviruses), episomal plasmids (e.g. F factor) and transposable genetic elements (e.g. P-elements, Ty-elements). These may be mutagenic due to their insertion and interruption of genes, or they may carry genes/regulatory elements affecting endogenous gene expression. Dispersed copies may facilitate illegitimate recombination.

rDNA Protocol

The entire protocol of rDNA experiments is to alter the genetic information contained within a DNA molecule. Considering the broad concept of mutation, rDNA Technology can be said to be mutagenic force.

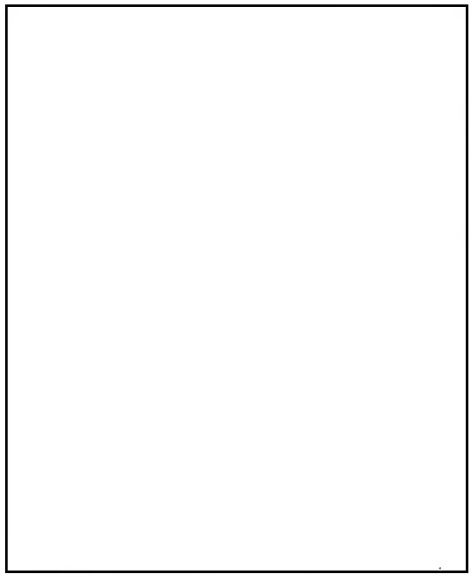
Importance of mutations

Mutations are both the sustainer of life and the cause of great suffering as shown below.

1. Mutation is the source of all genetic variation, the raw material of evolution. Without mutations and the variation that they generate, organisms could not adapt to changing environments and would risk extinction.
2. On the other hand, many mutations have detrimental effects, and mutation is the source of many human diseases (including many types of cancers).
3. Mutations also generate variety to be used in crop and livestock breeding programmes. Genetic crosses are meaningless if all in members of a species are identically homozygous for the same alleles.
4. Mutations serve as important tools of genetic analysis; such as Gregor Mendel's use of carefully selected variants of the garden pea; or Thomas Hunt Morgan discovering principles of mapping analyzing mutant fruit flies or determining the gene function etc.
5. Many biochemical processes are elucidated by using the knock-out mutants, such as the flowering pathway in angiosperms. Geneticists have unraveled the molecular details of development by studying mutations that interrupt various embryonic stages in *Drosophila*.

Direct PDF https://t.me/CivilServices_UPSC

https://t.me/Upsc_Pdf



upscpdf.com

Direct PDF 👉 https://t.me/CivilServices_UPSC